Users' behaviour patterns in academic libraries' OPACs: a multivariate statistical analysis

Aphrodite Malliari

*Department of Library Science and Information Systems,
Technological Educational Institution of Thessaloniki, Thessaloniki, Greece, and*

Daphne Kyriaki-Manessi

*Department of Library Science and Information Systems,
Technological Educational Institution of Athens, Athens, Greece*

Abstract

Purpose – This paper aims to present real time user searches in a Greek academic library OPAC (University of Macedonia Library) in relation to user profile.

Design/methodology/approach – Using as a test bed a Greek academic library and its OPAC's transaction logs along with a system implanted questionnaire, data were gathered, processed and analyzed using multivariate statistical analysis techniques.

Findings – In making a synthesis of the analyzed data, a series of questions related to everyday library work were answered, giving libraries a tool to apply the gained knowledge in order to make decisions regarding their OPAC, their user education programs and their reference services.

Research limitations/implications – The present paper focuses on the analysis of those variables that were considered to be the most representative for constructing a user profile.

Originality/value – This paper builds upon the techniques of data collection and presents a new tool for analyzing them statistically. Data derived from libraries were processed and analyzed statistically using the classical descriptive statistics. The suggested multivariate statistical method is designed to become a tool for analyzing qualitative data and to be used in a variety of library applications. It is also particularly helpful in analyzing cross-tabular data in the form of numerical frequencies and allows all associations amongst pairs of variables to be analyzed as well as each association between a variable and itself.

Keywords Online catalogues, User studies, Reference services, Academic libraries, Statistics, Greece

Paper type Research paper

Introduction

Every library's scope – as a service organization – is to provide quality services to its users and to strive continuously for their amelioration. Libraries, having stated their intention of satisfying their users, realized that their first task is to be informed about their needs and problems. Users do not always know what they want or are even able to explain which problem they have, what kind of difficulties they face when coming into a library or searching its online public access catalogue (OPAC). In order for these problems to be identified, examined and finally solved, librarians have developed user studies that have been carried out in many libraries all over the world. The main target of every user study is to bring to light the users' perspectives. Exploring users' behaviour, measuring users' satisfaction, meeting users' needs, even before they are

expressed, has become an art and a science in a library's quality services. User studies have become the medium for all.

In a nutshell, user studies are a mean of data collection – an old evaluation method for services provided (Siatri, 1999) – and a mode to re-establish the services, the needs and the priorities of every library. These are the tools for making decisions.

Taking into consideration this constant need for satisfied users, a user study was designed and implemented in the Library of the University of Macedonia in Thessaloniki. This academic library was selected because the University of Macedonia is a newly established institution, which hosts departments in the area of economic and social sciences and provides courses at both undergraduate and postgraduate levels. In that sense, the sample could be controlled (a population with similar areas of studies and interests) and at the same time provide ample data for statistical analysis. The Library is considered as a medium-sized academic library mainly because of its electronic resources. The University of Macedonia Library is electronically supported by the Horizon Information Portal of Dynix. Through the Library's OPAC, users have access to the book collection, which consists of approximately 50,000 volumes, to the collection of 1,168 foreign and Greek scientific journals, and also to collections such as the Law Introductory Essays, the Financial Brunch Studies (both available in Greek only), student theses of the University of Macedonia (available in Greek only) and finally to a collection of electronic books and journals.

The research set out to examine users' behaviour when searching the Library's online public access catalogue (OPAC), to identify the users' profiles and study their patterns, and also to measure their trials, errors and successes and to identify their needs and level of satisfaction. In order to assess these, a series of questions was formulated which became the parameters of the study and shaped the variables upon which the multivariate statistical analysis was based.

These questions are:

- Do users prefer the known-item searches – these being a straightforward author or title search – or do they venture for subject-item searches (not known-item searches)?
- Do the users' academic levels influence the effectiveness of their searches?
- Does the user's academic department (which means his scientific interests) influence the type and the effectiveness of the searches?
- Do users use all or even most of the system's capabilities for searching?
- Do users retrieve what they are looking for using the OPAC and, if not, what is the reason for their failure?

Methodology

Within this framework, this paper builds upon the techniques of data collection and presents a new tool for analyzing them. The data derive from the OPAC's transaction logs while users are accessing it. Their analysis focuses on the ability to apply multivariate statistical analysis on both qualitative and quantitative data and to derive valuable conclusions for locating users' behaviour patterns, failures and attributes that will eventually be applied by libraries to their OPAC software, their user training programs, etc. To date, the data derived from libraries were processed and analyzed statistically using univariate or even bivariate statistical techniques (Hardy, 1989;

Naylor and Walsh, 1994; Todd, 1994; Mantzopoulos, 1995; Healey, 1996; Smith, 1996; Stephen and Hornby, 1996; Hafner, 1998; Li, 1999; Ramsdale and Fuegi, 1999; Bertot *et al.*, 2000; Pors, 2000; Liu, 2001; Ambrožic, 2003). The suggested multivariate statistical method is designed to become a tool for analyzing statistically qualitative data and to be used in other library applications apart from the user behaviour analysis presented here. Data regarding usage of electronic sources, repositories, open access services versus subscription information sources can be some of the applications of the statistical tool presented in this study. It is also particularly helpful in analyzing cross-tabular data in the form of numerical frequencies, and allows all associations amongst pairs of variables to be analyzed as well as each association between a variable and itself.

Data collection and analysis

Data were collected by using both the transaction logs of the system and a questionnaire attached to the OPAC at the entry point. The University of Macedonia Library provided both the transaction logs and the technical support to attach on its OPAC the questionnaire, in return for the study' s results and recommendations.

Specifically, the transaction logs files provided a record of all different search types that the users have carried out on the system along with the systems' responses in chronological order. Apart from the precise and detailed recording of all search attempts and activities, the transaction logs had another advantage – they provided information about users' behaviour when searching the OPAC without the user being influenced by the researcher's presence. This guaranteed unbiased research data collection. The information gathered reflects the users' pure reactions to the system, their knowledge and ability to use it, and to what extent, and also the problems and failures they experienced. Through the transaction logs we managed to harvest real user reactions in real time and in real environment and, at the same time, anonymity was secured (as there is no user identification required for the use of OPACs) so the users' personal data were protected.

As McGlamery (1997) pointed out:

> . . . librarians have used transaction log analysis (TLA) since 1967. Log analysis at first used to determine system performance. In the late 1970s interest in user behaviour and performance on OPACs began to make use of the unobtrusive nature of TLA [. . .] More recently, work has been done to mine the transaction logs of Web servers and to utilize that analysis for library management information.

OPAC-centred user studies have been very common in libraries in all parts of the world and in all types of libraries. The most common are those focusing in identifying the main characteristics of specific groups in order to assist the library in the selection process of the most appropriate OPAC, the most user-friendly or the most sophisticated according to the groups' specifications (Burton and Hawkins, 1993; Park, 1997). Others were looking for answers to questions, such as what the users believe they want and what they really need (Wallace, 1993; Novotny, 2004). There have been also studies of users' attitudes and behaviour before and after the implementation of an online catalogue in their library (Martinez-Arellano, 1996; Dentinger, 1998; Lombardo and Condic, 2000). Furthermore, there were comparative studies of user behaviour over an OPAC and a card catalogue (Sridhar, 2004). Expressed concisely, it can be said that

user studies were used as a means to buttress collection development policies and as a decision making tool (Peters, 1996; Macewan, 1999; Banks, 2000).

Questionnaires, interviews, observation techniques and transaction logs were used for collecting data in user studies. Questionnaires, interviews and observation techniques could give user profiles, user behaviour and attitudes. At the same time, they require the researchers' presence, which could influence the results. In contrast, transaction logs can give unbiased data but no user profile. The paper by Peters (1996) is a very good starting point for understanding the advantages and disadvantages of transaction log analysis. For the present study, a combination of techniques was devised so as to be able to compile a precise user profile without interfering with the user's activities and at the same time to allow us to crop the benefits of the transaction logs. A very brief online questionnaire was inserted in the system's prompt. The questionnaire consisted of five closed-type questions and focused primarily on the user's affiliation and status in the university; the department that he/she is enrolled or teaches at and the level of studying or professorship.

The aforementioned online questionnaire consisted of the following questions that the users had to answer before entering the OPAC:

- type of search (Simple, Advanced);
- user category (Undergraduate, Graduate, Faculty, or "other");
- department (the 11 departments of the University, or "other")
- year of enrolment (first, second, third, fourth, degree, other); and
- faculty rank (Lecturer, Assistant Professor, Associate Professor, Professor, other).

The user had the choice not to participate in the research by simply answering "other" to the questions and not giving any of the profiling data requested. It is also evident that anonymity was preserved as all entries to the system do not require a user name and use of the OPAC is open to all members of the academic community.

Regarding the type of search, it is up to the user to decide whether to undertake a simple or an advanced search without having to follow guidelines or having help messages available. It is easy to choose one type of search and then turn to another.

Data were collected throughout the academic year (from early November 2001 to June 2003, which are the full months of two academic semesters in Greek universities) and include peak times such as examination periods, semester working periods including final papers, etc., and low-use periods such as Christmas and Easter vacations. Then the data were studied, processed and analyzed statistically.

From the collected data a sample was formulated containing data deriving from one week for every month of the study. The sample contained 4,287 sessions (log-ins). Only 43 percent of the log-in sessions contained searches, which produced 1,837 searching sessions. These 1,837 sessions were processed and analyzed statistically. It should be noted that a great number of connections (57 percent, or 2,450 log-ins) to the Library's server were made without the users actually performing a search. For example, during the period of November 26-December 2, 533 connections were made to the library's server. Of these 533 connections only 307, i.e. 58 percent, were made with the purpose of searching the OPAC. This is presented in Figure 1. This is mainly due to a system error that kept logging the users out and prompting a message text recorded in the

transaction log files: "Message text: Socket is not connected". This message was frequently recorded in December's transaction logs. Also, some of the visits were made simply with the sole purpose of explore the Library's site and with no intention to use the OPAC.

In total, during the whole study period, the transaction logs recorded 17,067 sessions. All sessions were divided into three groups. The first group contains the sessions made from users who answered the questionnaire thus giving us their profile. The second group contains those sessions that were made from users who chose the option "other" in every question, thus bypassing the questionnaire and providing us with no profile. Finally, the third group contains users who, connecting with the OPAC, did not come across the questionnaire at all. There were some cases in which users could enter the OPAC without coming across the questionnaire. Such cases include users having the library's OPAC as a bookmark in their PC before the implanting of the questionnaire.

For the purposes of this study and the application of multivariate statistical analysis the last two groups presented no interest and consequently were excluded from the analysis. This is due to the fact that these two groups by not supplying the questionnaire could not provide us with user profiles; therefore no correlation could be made to the recorded data on the transaction logs. Qualitative characteristics could not be provided solely through the transaction logs data without the combination of questionnaires. Consequently, the study focused on data deriving from the first group of users and their approximate 754 search sessions.

Data deriving from the transaction logs
Transaction logs contain data from all users' activities from the moment they enter the system to the moment they either log off, or the system fails them or for any other system the connection is interrupted. Users get connected to the Library's OPAC either from a terminal within the Library or from an access point within the University's campus or from their personal PC. A web browser such as Netscape Navigator or Microsoft Internet Explorer is used. The session's duration depends on the user's needs, on the systems response and capability to carry out the session without interruptions. If the user, when ending their search does not log out, the system gives a 20-minute period and then automatically logs the user out.

Transaction logs files keep information about the date, the time, the duration and the IP address of the PC. A registry number is assigned automatically by the system the moment a session starts. This makes every session distinct. Within every session
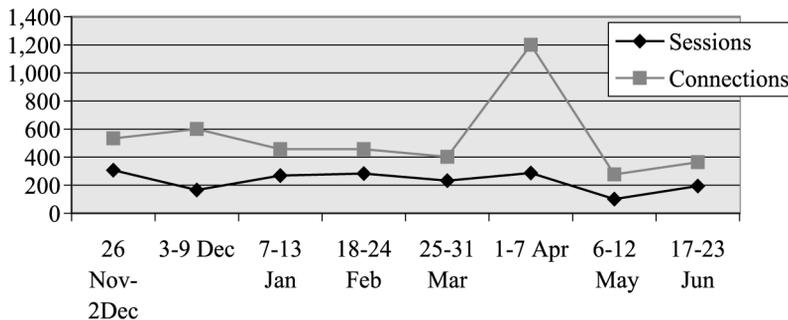

Figure 1.
Connections and sessions
in OPAC during the period
November-June 2002

all search activities are recorded. These are the typical transaction log files contents. In addition to these, the questionnaire and the user's answers were also recorded as part of the sessions recorded on the transaction logs.

Data analysis

Figure 2 presents a sample of a transaction log file. The user of the particular session is an Associate Professor in the Department of Balkan Studies, as shown by "USMARC_Simple_en_dep_bal_oth_assoc" (questionnaire). He/she is searching the catalogue by entering the term "counselling skills" first in the title field. Secondly, he/she prefers to browse the catalogue using the subject heading "Counselling – Research – Methodology", a subject heading that is related but far from useful. As the user performs another search, this time using the terms "counselling skills" in the subject field, it is assumed that the results did not satisfy him/her. Finally, he/she makes two more attempts without results and abandons. Thus, as the user has not made any selection, it is assumed that the system produced no relevant results.

By making a selection, the user chooses to see one or more of the hit records either in their full display or in MARC format.

This is just an example – the analysis of the full amount of data collected and the correlation with the user profiles emerging from the questionnaires helped us to formulate the variables based upon which qualitative analysis was built.

The variables presenting user category, department and search type were chosen as the most representative variables for constructing a user profile. As the data analysis progressed, it became evident that users had one characteristic, for example students of the Computing Department preferred author searches, whilst students of the International Studies Department preferred subject searches. Teaching staff, especially those of lower academic rank (lecturers), performed more searches (in terms of numbers) and focused on subject searches rather than author or title ones (known-item searches).

In addition, the following variables were taken into account and analyzed the same way:

- number of selections to the search outcome;
- mistakes made;
- terms that were repetitively re-entered in searches during the same session; and
- all of the above in relation to user category and to search type.

The "user category" variable

Figure 3 presents the analysis of data regarding user category. As can be seen, 58 percent of OPAC users are graduate students, 25 percent are undergraduate students, 11 percent are teaching faculty and the remaining 6 percent belong to the category characterized as "other", which refers to University research staff. This reflects the fact that the graduate population at the University is the largest user group, but it also reflects a behaviour pattern that characterizes the specific group that it is geared to research, and the fact that their level of studies requires a much heavier use of the library than the undergraduate level. The faculty's rather small representation also reflects their smaller population number in relation to students.

Studying the part of the Burt table (see Table I), where the variable of the user category crosses with the variables of selections, mistakes and terms re-entered, the

2-05-29 18:06:40 C:\Programs\WebPAC1.35\wgbroker.exe 270 INFO: new session for client: 62.1.6.7
2-05-29 18:06:40 C:\Programs\WebPAC1.35\wgbroker.exe 270 INFO: run key: 200205291806400638270
2-05-29 18:06:41 .\wgsubmit.exe 268 INFO: run: 200205291806400638270 port: 1557 client: 62.1.6.7
2-05-29 18:06:41 .\wgsubmit.exe 268 INFO: entering command loop
2-05-29 18:06:42 C:\Programs\WebPAC1.35\wgbroker.exe 270 INFO: started wgsubmit
2-05-29 18:06:42 .\wgsubmit.exe 268 INFO: entering PROCESSING state
2-05-29 18:06:42 .\wgsubmit.exe 268 INFO: processing command: -access
top.USMARC_Simple_en_dep_bal_oth_assoc
2-05-29 18:06:42 .\wgsubmit.exe 268 INFO: run 200205291806400638270 session 1 initially connected to
database: USMARC_Horizon_en
2-05-29 18:06:42 .\wgsubmit.exe 268 INFO: run 200205291806400638270 session 1 database
USMARC_Horizon_en command(search)
2-05-29 18:06:43 .\wgsubmit.exe 268 ERROR: Invalid qualifier: UK in node:
top.USMARC_Simple_en_dep_bal_oth_assoc
2-05-29 18:06:43 .\wgsubmit.exe 268 INFO: entering ACCEPTING state
2-05-29 18:07:02 .\wgsubmit.exe 268 INFO: entering PROCESSING state
2-05-29 18:07:02 .\wgsubmit.exe 268 INFO: processing command: 1 scan open TA counselling skills
2-05-29 18:07:02 .\wgsubmit.exe 268 INFO: run 200205291806400638270 session 1 database
USMARC_Horizon_en command(scan open TA counselling skills)
2-05-29 18:07:20 .\wgsubmit.exe 268 INFO: entering ACCEPTING state
2-05-29 18:07:41 .\wgsubmit.exe 268 INFO: entering PROCESSING state
2-05-29 18:07:41 .\wgsubmit.exe 268 INFO: processing command: 1 search open TK counselling skills
2-05-29 18:07:41 .\wgsubmit.exe 268 INFO: run 200205291806400638270 session 1 database
USMARC_Horizon_en command(search open TK counselling skills)
2-05-29 18:07:49 .\wgsubmit.exe 268 INFO: entering ACCEPTING state
2-05-29 18:08:13 .\wgsubmit.exe 268 INFO: entering PROCESSING state
2-05-29 18:08:13 .\wgsubmit.exe 268 INFO: processing command: 1 search -redir 2-0 open R_SUBJECT
Counseling -- Research -- Methodology.
2-05-29 18:08:13 .\wgsubmit.exe 268 INFO: run 200205291806400638270 session 1 database
USMARC_Horizon_en command(search -redir 2-0 open R_SUBJECT Counseling -- Research --
Methodology.)
2-05-29 18:08:18 .\wgsubmit.exe 268 INFO: entering ACCEPTING state
2-05-29 18:08:49 .\wgsubmit.exe 268 INFO: entering PROCESSING state
2-05-29 18:08:49 .\wgsubmit.exe 268 INFO: processing command: 1 search open SK counselling skills
2-05-29 18:08:49 .\wgsubmit.exe 268 INFO: run 200205291806400638270 session 1 database
USMARC_Horizon_en command(search open SK counselling skills)
2-05-29 18:08:50 .\wgsubmit.exe 268 INFO: entering ACCEPTING state
2-05-29 18:09:03 .\wgsubmit.exe 268 INFO: entering PROCESSING state
2-05-29 18:09:03 .\wgsubmit.exe 268 INFO: processing command: 1 search open SK counselling
2-05-29 18:09:03 .\wgsubmit.exe 268 INFO: run 200205291806400638270 session 1 database
USMARC_Horizon_en command(search open SK counselling)
2-05-29 18:09:04 .\wgsubmit.exe 268 INFO: entering ACCEPTING state
2-05-29 18:09:14 .\wgsubmit.exe 268 INFO: entering PROCESSING state
2-05-29 18:09:14 .\wgsubmit.exe 268 INFO: processing command: 1 search open SK counseling
2-05-29 18:09:14 .\wgsubmit.exe 268 INFO: run 200205291806400638270 session 1 database
USMARC_Horizon_en command(search open SK counseling)
2-05-29 18:09:28 .\wgsubmit.exe 268 INFO: entering ACCEPTING state
2-05-29 18:29:28 .\wgsubmit.exe 268 INFO: leaving command loop, rc is: 2
2-05-29 18:29:28 .\wgsubmit.exe 268 INFO: wgsubmit run: 200205291806400638270 is exiting

Figure 2.
A users' session from the
transaction logs file

following results were rendered. Fifty-two percent of the graduate students repeat
more than one time by entering the same terms in repetitive searches in the OPAC.
Thirty-four percent make mistakes while searching (these mistakes can be from
typographical errors and spelling to misuse of the system, etc.). For
undergraduate/PhD students, these rates are reduced to 42 percent for re-entering
terms and to 29 percent regarding mistakes. In contrast, 37 percent of the teaching
faculty re-enter the same terms in repetitive searches and 19 percent make mistakes.

Moreover, we noticed that graduate students exit the system without making any selection more frequently than teaching faculty members (33 percent of graduate students to 29 percent of faculty members).

In summary, it can be said that the quality of searches is directly dependent upon the user's experience and familiarity with the OPAC.

The "department" variable

One of the questions that this study attempted to answer was whether belonging to a particular University department in any way affected the user's behaviour in searching the OPAC. In order to answer this question, the representation of each department in the data collected was evaluated first. Figure 4 depicts these numbers.

Specifically:

- 13 percent of the OPAC's users are members of the Economics Department;

- 12 percent are from the Department of Business Administration;

- 13 percent are members of the Department of International & European Economic & Political Studies;

- 10 percent are members of the three new departments, these being Music Studies, Balkan & Slavic Studies, and Educational & Social Policy;

- a further 9 percent are members of the Applied Computer Science Department;

- another 25 percent are members of the three departments of master studies, these being Information Systems, Business Administration and Economic Science; and

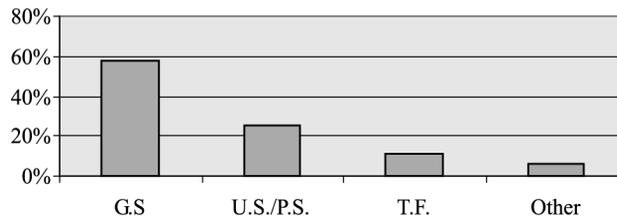- the remaining 11 percent are faculty members of all departments.

Figure 3.
User category

	Graduate	Undergraduate	Faculty	Other
0 selections	32.9	29.8	28.9	29.2
1 selection	21.6	23.4	26.5	20.8
2 selections	14.9	14.4	18.1	10.4
3-4 selections	12.9	18.1	13.3	22.9
≥5 selections	17.7	14.4	13.3	16.7
Same words	51.7	41.5	37.3	62.5
New words	48.3	58.5	62.7	37.5
Without mistakes	66	71.3	80.7	72.9
With mistakes	34	28.7	19.3	27.1

Table I.
Relation of user category with selection, re-entering terms and mistakes

Note: Figures given are percentages

30%
25%
20%
15%
10%
5%
0%

Economic St. | Business Adm. | International St. | New Dep. | Applied Computer St. | Finance | MSc | Other
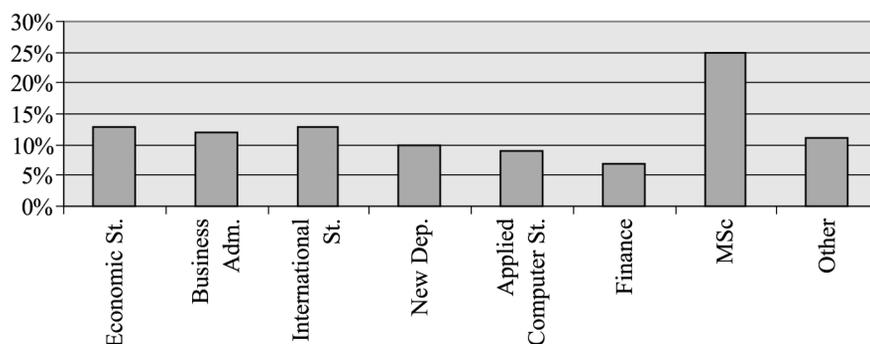
Figure 4.
University's departments

To clarify this, faculty members were classified, for the purposes of this study, in the category "other" in order to avoid internal frictions.

The aforementioned percentages indicate that the variable "department" does not produce any significant disparities that would allow conclusions to be drawn that disciplinary orientation affects the use of the academic library. For this reason no further analysis was made regarding the variables of selection, mistakes and terms re-entered.

Search type
The analysis of this variable presents for the library the most important factor in order to assess the way its OPAC is used. Figure 5 presents the data collected from the transaction logs.

Specifically, 41 percent of the Library's users prefer the title for their first search in the OPAC. When the user chooses this type of search, he/she could either search the catalogue by entering keyword(s) from the title or browse the catalogue alphabetically by entering the first word(s) from the title. Twenty-seven percent of users prefer to search the catalogue using a keyword from the author's name or the exact name, 23 percent prefer the subject search type and the remaining 8 percent search the OPAC via the types "basic keyword", "series" and "ISBN/ISSN". When the user prefers the subject search type, he/she enters a term and decides to search this term either as a keyword or as the exact subject heading in alphabetical order from the subject headings (a controlled language is used for subject headings). When the user enters a term in the field of "basic keyword", the system searches everywhere in the catalogue.

In analysing the data further, it can be seen that from all users preferring the title search, 31 percent do not make any selection from the results produced, whereas from
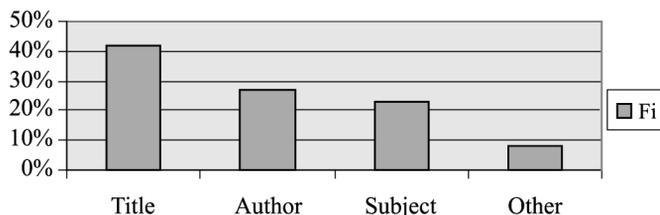

50%
40%
30%
20%
10%
0%

Title | Author | Subject | Other

☐ Fi

Figure 5.
Search type

116

those who perform an author search, the selection rate is higher, as only 16 percent do not make any selection from the results produced.

Contrary to this, of those users who have searched the catalogue for the first time using a subject search or other types of searches, such as basic keyword, series and ISBN/ISSN, a good 43 percent and 50 percent, respectively decide not to make any selection from the results produced (see Table II).

This can be explained from the simple fact that users searching via an author are actually performing a known-item search, meaning that they know beforehand what they are looking for, and as soon as they locate it they select it. Users performing a title search fall under the same category, but, as titles tend to have more words, the accuracy level diminishes and the known-item search sometimes fails – hence the higher level of users not proceeding to a selection. The high percentage of users that do not make any selection after performing a subject search in the OPAC can be attributed to series of factors. Subject searches are characterized as unknown-item searches, which means that the user does not know if a certain item exists but he/she is looking for what is available in a particular subject area. This type of search requires some basic searching skills in terms of terminology, using some knowledge of the search engine and the basic techniques if he/she were to combine subjects. It also presumes that the user feels confident enough to use the subject feature of the system, that he/she knows the capabilities of the system and that the library is accommodating those needs with widely known terminology and approved standards. This latter seems to be a factor explaining the overall low rate of subject searches and the even lower rate regarding the selection from the retrieved data. It basically suggests that practically half of the users performing an unknown item search do not get the desired results.

On looking closely at the title searches, it can be seen that a further 53 percent try new terms in their repetitive searches, while 71 percent do not make any kind of mistakes. Of those users who prefer the author searches, 61 percent enter new versions or names and 69 percent do not make mistakes. Accordingly, 61 percent of the "subject-users" re-enter the same terms in repetitive searches (Table II) while the percentage of mistakes is 32 percent. The high rate of users re-entering again and again the same words when they are searching the OPAC by subject agrees with the high rate of users making no selections from the search results. It is obvious that the users insist on searching the OPAC without being aware of the terminology used or the alternative synonyms through the use of a thesaurus. We assume that the user is not

	Title	Author	Subject	Other
0 selections	31.1	16.2	43.2	50
1 selection	23.4	26.5	15.9	24.2
2 selections	14.1	21.1	11.9	6.5
3-4 selections	16.7	14.7	13.6	9.7
$\geq$5 selections	14.7	21.6	15.3	9.7
Some words	46.8	38.7	60.8	51.6
New words	53.2	61.3	39.2	48.4
Without mistakes	71.5	68.6	68.2	64.5.
With mistakes	28.5	31.4	31.8	35.5

Table II.
Relation of search type with no of selections, re-entering and mistakes

Note: Figures given are percentages

able to recognize the fact that he/she is using the "wrong" keywords or phrases or to go a step further and replace them with alternatives.

Users' behaviour patterns

117

Data synthesis and conclusion

In making a synthesis of the analyzed data, a series of questions was derived that can be answered based on the correlation of the data analyzed above. The questions were formulated in order to give libraries a tool to apply the knowledge gained from the analysis to everyday library work, and to help them make decisions regarding their OPAC, their user education programs and their reference services.

Do users prefer known-item or subject (unknown-item) searches?

The users of the library clearly prefer "known-item" searches, those searches that were made using the title or the author of a document. They probably consider this type of search easier and "safer" as fewer mistakes are made and they usually render more selections to results. The reasons for avoiding "subject-item" searches are mainly the users' ignorance of the terminology used (subject headings), their weakness in expressing the subject with the proper keywords, and their lack of knowledge of the system's capabilities. Last but not least, the fact that the system has as default choice of the search type "title" plays an important role. Many studies have shown that users do not easily change the default choice, except for those who are considered "experienced" and consciously select another type of search (Jones *et al.*, 2000).

Does the academic level of users affect the effectiveness of the searches?

In analyzing the data gathered from the transaction logs in relation to the data derived through the questionnaire, it became obvious that the user's academic level affects the effectiveness of their searches. It was observed that teaching faculty members prefer "known-item" searches and then move on to make selections from the systems response list. This can be explained by the fact that faculty members seem to have a clearer idea of what they are looking for. Their searches are based on experience, they also are far more accurate than the students, and the mistake rate is very low, as is their rate in re-entering the same words. It can be said that teaching faculty know exactly what they want and how to retrieve it. On the other hand, the students re-enter the same words, make mistakes and readily abandon the whole effort of retrieving information from the catalogue. Frequently, they do not proceed in any selection from the results produced. Graduate students, however, having more experience and being more conscious of the process, are more effective in their searches, have a lower rate of mistakes, and move on to the selection process quite effectively.

Does the user's academic department (which means his/her scientific interests) influence the type and effectiveness of the search?

Data analysis could not provide an answer to the above question. This is due to the fact that every department had almost the same weight as a variable. Thus, it was concluded that scientific interest and field of study do not influence either the success of the search or the chosen search type. The only exception was the increased use of the OPAC from the fourth-year students of the Department of International & European Economics & Political Studies. This student group presented a heavier use of the OPAC as 36 percent of the fourth-year students came from this department and the rest

were dispersed to the other departments. A closer look at their curriculum and the syllabi of the specific year's courses provided an explanation – some of their courses relied heavily on term papers rather than exams and the students had to carry out research projects as well. What it means in this situation is that the scientific orientation of a department does not affect library use, but the departmental policy regarding term papers, assignments and encouragement of independent student research is the key to library use. Furthermore, it is obvious that the library, by reporting these findings to the departments, could act as a barometer for teaching methods and at the same time could set the parameters for cooperation.

Do users make the most of the system's potentials for searching?
Even from the early stages of data analysis it became quite obvious that users do not employ all of the system's potential for searching. Specifically, only 2 percent of the users combine terms using Boolean operators. The percentage (0.1 percent) of users that have refined their search by using the publication date is also strikingly low. Also, only 1 percent of users utilize the option of collection limitations. It is evident that users do not know the system's capabilities and/or they hesitate to use it – the need for user education and user support systems is clear. These rates also might indicate – and this should be studied further – that the system is not sufficiently "user friendly", that it does not readily provide the search options, and that its layout and screens are not set out properly.

An additional piece of information regarding the search type ISBN and ISSN indicates another low use. Only 4 percent of searches are made via this type. However, this is to be expected. It is the "known item" search that the user cannot easily remember – it requires notes and precision, and other searches of the same type such as author or title can offer the same results. ISBN searches are usually performed by information professionals and as such are used by the library staff and not by the library users.

Do users retrieve the desired information? And if not, why not?
Examining the data deriving from the transaction logs, and specifically the search types and their success rate (meaning the retrieval of the acquired information and the consecutive selection of items from the list) a higher failure rate was observed in the search types of:

- subjects;
- ISBN;
- series; and
- basic keyword.

Failure in searching by subject is mainly due to the user's unfamiliarity with the library's terminology. Usually, the user searches using plain, everyday terms to express research concepts – in some cases he/she uses the field terminology. However, the university library uses the controlled vocabulary of subject headings formulated by reversing the main concept, which usually follows a descriptive identifier. This artificial language very often includes terminology that the user is highly unlikely to use or even think of as a possibility. Subject headings often derive from deliberate

translations from the Library of Congress (LC) equivalents not always used in Greek, thus creating a daedal of words and ill-served concepts. As a result, the terms in natural language expressed by users do not match those assigned by librarians in the controlled vocabulary. This is a clear message to the library that simplification and standardization of the vocabulary used must become a top priority, along with user education.

Furthermore, on looking at the ISBN mistakes it was noticed that, instead of entering the ISBN, many mistakes recorded the entering of other numbers appearing on an item, such as the LC number, the Dewey Decimal Classification (DDC) number or even whole phrases or keywords instead of numbers. This clearly indicates the need for user education. Series searches indicate flows such as using keywords, titles or even authors instead of a series attribute. The same holds for keyword searches, which were mostly mistaken for subject searches. Overall the aforementioned failures indicate some of the library's weaknesses in providing its users with a standardized vocabulary in natural language and also the great needs of users for education in library use and usage. The failures also provide a good indication of what should be incorporated in user education sessions.

Suggestions

Based on the aforementioned observations, the Library should re-examine specific issues in its services, its user education programs and update its software. The Library has to amend its services, traditional and electronic, in order to satisfy users and to meet their information needs. Changes and improvements have to be made both in bibliographic instruction and in the functionality of the OPAC applications, as bibliographic instruction is supplementary to the proper use of the OPAC by the Library's users and *vice versa*. Therefore, the Library has to provide users with continuing education, and especially bibliographic instruction (BI) programs or online information literacy programs, in order that they can learn how to use and retrieve information from the catalogue and the other information sources. The need for constant BI is referred to in many OPAC studies (Peters, 1989; Hunter, 1991; Ciliberti *et al.*, 1998) in order to address the problems rising from wrong or inappropriate use of online catalogues. In addition to these programs, the Library should offer on-site help to new users during the fall semester and also establish a help desk. On the other hand, the online catalogue system's administrators have to examine and perhaps change the default search strategy that the system prompts to users, implement an automated correction of spelling mistakes, insert a prompt in faulty usage that will guide the user to correct his/her search (for example to stop them inserting words or phrases in numeric searches such as ISBN), and also an online help facility with short and precise answers with examples on how to use the catalogue. Blecic *et al.* (1998, 1999), having analysed OPAC users, revealed that the problems users had experienced while searching the OPAC were addressed by changing and simplifying the OPAC's introductory screens. As far as subject searches failures are concerned, it is suggested that, apart from the aforementioned bibliographic instruction, the OPAC's software should be re-established so that terms in natural language that are more familiar to users could meet those of the controlled vocabulary of subject headings. As Drabenstoot and Weller (1996) report, "These problems [referring to problems

experienced by users with subject searches] are indicative of the need for a new design for subject access to online catalogs".

In summary, this attempt to find out users' behaviour regarding the OPAC will help the Library to re-establish its policy, always with the target in mind of ameliorating as well as expanding its services. It is evident that Library's policy could be influenced both in terms of the anthropocentric character of its software and in its efficiency and effectiveness.

References

Ambrožič, M. (2003), "A few countries measure impact and outcomes: most would like to measure at least something", *Performance Measurement and Metrics*, Vol. 4 No. 2, pp. 64-78.

Banks, J. (2000), "Are transaction logs useful? A ten year study", *Journal of Southern Academic and Special Librarianship*, Vol. 1 No. 3, available at: http://southernlibrarianship.icaap.org/content/v01n03/banks_j01.html

Bertot, J.C., McClure, Ch.R. and Joe, R. (2000), "Developing national network statistics and performance measures for US public libraries: issues, findings and recommendations", *Performance Measurement and Metrics*, Vol. 1 No. 1, pp. 15-41.

Blecic, D., Dorsch, J., Koenig, M. and Bangalore, N. (1999), "A longitudinal study of the effects of OPAC screen changes on searching behaviour and searcher success", *College & Research Libraries*, Vol. 60 No. 6, pp. 515-30.

Blecic, D., Bangalore, N., Dorsch, J., Henderson, C., Koenig, M. and Weller, A. (1998), "Using transaction log analysis to improve OPAC retrieval results", *College & Research Libraries*, Vol. 59 No. 1, pp. 39-50.

Burton, P.A. and Hawkins, A.M. (1993), "Attitudes to an online public access catalogue in an academic library", *Library Management*, Vol. 14 No. 5, pp. 13-15.

Ciliberti, A., Radford, M.L., Radford, G.P. and Ballard, T. (1998), "Empty handed? A material availability study and transaction log analysis verification", *The Journal of Academic Librarianship*, Vol. 24 No. 4, pp. 282-9.

Dentinger, S. (1998), "The public electronic library: web-menuing versus OPAC cataloguing", *Library Hi Tech*, Vol. 16 Nos 3/4, pp. 89-94.

Drabenstoot, K.M. and Weller, M.S. (1996), "Failure analysis of subject searches in a test of a new design for subject access to online catalogs", *Journal of the American Society for Information Science*, Vol. 47 No. 7, pp. 519-37.

Hafner, A.W. (1998), *Descriptive Statistical Techniques for Librarians*, American Library Association, Chicago, IL.

Hardy, E.D. (Ed.) (1989), *Statistics for Managing Library Acquisitions*, American Library Association, Chicago, IL.

Healey, J.F. (1996), *Statistics: A Tool for Social Research*, Wadsworth, Belmont, CA.

Hunter, R.N. (1991), "Successes and failures of patrons searching the online catalog at a large academic library: a transaction log analysis", *RQ*, Vol. 30, pp. 395-402.

Jones, S., Cunningham, S. and McNab, R. (2000), "A transaction log analysis of a digital library", *International Journal on Digital Libraries*, No. 3, pp. 152-69.

Li, X. (1999), "Library web page usage: a statistical analysis", *The Bottom Line: Managing Library Finances*, Vol. 12 No. 4, pp. 153-9.

Liu, Y.Q. (2001), "The use of statistics in the managerial process by public library directors in China and the USA: a cross-country survey", *Library Management*, Vol. 22 Nos 4/5, pp. 187-94.

Lombardo, Sh.V. and Condic, K.S. (2000), "Empowering users with a new online catalog", *Library Hi Tech*, Vol. 18 No. 2, pp. 130-41.

McGlamery, P. (1997), "MAGIC transaction logs as measures of access, use and community", *The Journal of Academic Librarianship*, Vol. 23 No. 6, pp. 505-10.

Macewan, B. (1999), "Understanding users' needs and making collections choices", *Library Collections, Acquisitions and Technical Services*, Vol. 23 No. 3, pp. 315-20.

Mantzopoulos, V.I. (1995), *Statistics for Social Sciences*, Prentice Hall, Englewood Cliffs, NJ.

Martinez-Arellano, F. (1996), "User attitudes and behaviour before and after the implementation of an on-line catalogue", *International Information & Library Review*, No. 28, pp. 275-87.

Naylor, M. and Walsh, K. (1994), "A time-series model for academic library data using intervention analysis", *Library & Information Science Research*, Vol. 16 No. 4, pp. 299-314.

Novotny, E. (2004), "I don't think I click: a protocol analysis study of use of a library online catalog in the internet age", *College and Research Libraries*, Vol. 65 No. 6, pp. 527-37.

Park, I. (1997), "A comparative study of major OPACs in selected academic libraries for developing countries: user study and subjective user evaluation", *International Information & Library Review*, Vol. 29 No. 1, pp. 67-83.

Peters, T.A. (1989), "When smart people fail: an analysis of the transaction log of an online public access catalog", *The Journal of Academic Librarianship*, Vol. 15 No. 5, pp. 267-73.

Peters, T.A. (1996), "Using transaction logs analysis for library management information", *Library Administration and Management*, Vol. 10 No. 1, pp. 20-5.

Pors, N.O. (2000), "Statistics and teaching in departments of library and information studies", *The Bottom Line: Managing Library Finances*, Vol. 13 No. 1, pp. 16-20.

Ramsdale, P. and Fuegi, D. (1999), "LIBECON2000 and the future of international library statistics", *The Bottom Line: Managing Library Finances*, Vol. 12 No. 4, pp. 160-2.

Siatri, R. (1999), "The evolution of user studies", *Libri*, No. 49, pp. 132-41.

Smith, M.L. (1996), "Using statistics to increase public library budgets", *The Bottom Line: Managing Library Finances*, Vol. 9 No. 3, pp. 4-13.

Sridhar, M.S. (2004), "OPAC vs card catalogue: a comparative study of user behaviour", *The Electronic Library*, Vol. 22 No. 2, pp. 175-83.

Stephen, P. and Hornby, S. (1996), *Simple Statistics for Library and Information Professionals*, Library Association Publishing, London.

Todd, R.J. (1994), "The collection of library statistics for supplier evaluation: a user-friendly guide", *Library Acquisitions: Practice & Theory*, Vol. 18 No. 1, pp. 61-6.

Wallace, P. (1993), "How do patrons search the online catalog when no one's looking?", *RQ*, Vol. 33 No. 2, pp. 239-52.

Further reading

Borgman, Ch.L. (1996), "Why are online catalogs still hard to use?", *Journal of the American Society for Information Science*, Vol. 47 No. 7, pp. 493-503.

Chen, H.-M. (2000), *An Analytical Approach to Deriving Usage Patterns in a Web-based Information System*, University of California, Information Management and Systems Graduate Division, Berkley, CA.

Jansen, B.J., Spink, A. and Saracevic, T. (2000), "Real life, real users, and real needs: a study and analysis of user queries on the web", *Information Processing & Management*, Vol. 36 No. 2, pp. 207-27.

Ke, H.-R., Kwakkelaar, R., Tai, L. and Chen, L. (2002), "Exploring behaviour of e-journals users in science and technology: transaction log analysis of Elsevier's Science Direct onsite in Taiwan", *Library & Information Science Research*, Vol. 24 No. 2, pp. 265-91.

Koenig, M.E.D. (2003), "Knowledge management, user education and librarianship", *Library Review*, Vol. 52 No. 1, pp. 10-17.

Mahoui, M. and Cunningham, S.J. (2000), "A comparative transaction log analysis of two computing collections", *Lecture Notes in Computer Science*, Vol. 1293, pp. 418-23.

Malliari, A. (2005), "Information retrieval from academic libraries' automated systems and statistical analysis", Department of Applied Informatics, University of Macedonia, Thessaloniki.

Moukdad, H. and Large, A. (2001), "Users' perceptions of the web as revealed transaction log analysis", *Online Information Review*, Vol. 25 No. 6, pp. 349-58.

Wyly, B.J. (1996), "From access points to materials: a transaction log analysis of access point value for online catalogue users", *Library Resources & Technical Services*, Vol. 40 No. 3, pp. 211-36.

About the authors
Aphrodite Malliari is a Reference Librarian at the Library of TEI of Thessaloniki, responsible for online bibliographic instruction, and an adjunct lecturer at the Department of Library and Information Systems of the Technological Educational Institution of Thessaloniki. Among her interests are patron needs, information retrieval, multivariate statistical analysis and academic libraries. Aphrodite Malliari is the corresponding author and can be contacted at: malliari@libd.teithe.gr

Daphne Kyriaki-Manessi is an Associate Professor at the Technological and Educational Institution of Athens at the Department of Library Science and Information Systems. Academic interests include knowledge management, structures of data and data mining as well as the field of information organizations and their setting within information society.