Int J Digit Libr (2010) 11:225–238
DOI 10.1007/s00799-011-0075-7

Understanding user requirements and preferences for a digital library Web portal

Maristella Agosti · Franco Crivellari · Giorgio Maria Di Nunzio · Silvia Gabrielli

Published online: 30 September 2011

Abstract This article reports the findings of a user study conducted in the context of the TELplus project to gain insights about user needs and preferences for the digital library services offered by The European Library Web portal. The user requirements collection for the Web portal was designed by adopting a comprehensive survey approach. This combined explicit user feedback with implicit usage data so as to provide a more in-depth analysis of user experience with the portal. The analysis conducted shed light on likely motivations for both participant usage and reluctance to use the services provided, leading to more informed decisions on how to refine, improve, and present Web portal services to their future users. The lessons learnt from this case study also contributed to the development of an integrated methodological framework which provided insights for the future design and evaluation of digital library Web portals and services.

Keywords User requirements for digital library Web portals · Preferences for Web portals · User survey · Log data · User interaction with digital contents

M. Agosti · F. Crivellari · G. M. Di Nunzio (✉)
Department of Information Engineering, University of Padua, Via Gradenigo, 6/a, 35131 Padova, PD, Italy
e-mail: giorgiomaria.dinunzio@unipd.it

M. Agosti
e-mail: maristella.agosti@unipd.it

F. Crivellari
e-mail: franco.crivellari@unipd.it

S. Gabrielli
Ubiquitous Interaction Group, CREATE-NET, Via Alla Cascata 56/D, 38123 Trento, TN, Italy
e-mail: silvia.gabrielli@create-net.org

1 Background

The design of digital library Web portals and services presents a major challenge in terms of first understanding and then matching the different, evolving needs and preferences of their target user groups. During the user-centred design of a prototype for The European Library (TEL)[1] Web portal, we approached this challenge by applying comprehensive survey methods based on a combination of both explicit and implicit measures of user interaction. The aim was to better uncover and also track user needs, interests and preferences for the system designed.

Evaluation of digital library services based on user explicit feedback can take many forms depending on the objectives of the analysis [14,26].

When a prototype of a Digital Library (DL) service is available, the effectiveness, efficiency, and/or satisfaction of the user experience with a particular interface is often carried out [23,25].

In [24], user satisfaction regarding reference services was investigated through surveys, observations and follow-up focus groups. Hill et al. [16] used multiple methods, including online surveys, ethnographic studies, focus groups, and user comments to incorporate user feedback in the ongoing design and implementation of the Alexandria DL at the University of California, USA.

In [12], a multi-disciplinary, human-centred perspective on the design of DL and the global information infrastructure is provided.

Any evaluation of DL collections and services must be inevitably taken into account the characteristics of the target

[1] http://www.theeuropeanlibrary.org/.

user groups. However, the study of user needs and preferences has often eluded long term, quantitative analysis [11]. Since, user preferences and choices tend to be highly transient and specific [18], both the collection of explicit measures regarding user satisfaction, interest and preference, and implicit measures regarding usage (e.g., based on retrieval patterns recorded for large communities of users) can provide valuable results. As an implicit measure of user behavior, log data analysis is certainly not intrusive. However, it is more demanding, compared to a questionnaire, when it comes to reconstructing and disambiguating what happened during a user session to correctly infer user motivations and preferences.

Although, the combination of different sources of evidence requires more effort in the data analysis process, this approach has received attention by an increasing number of researchers over the years. In [17], Ingwersen and Järvelin report that it seems more scientifically informative to combine logs together with observation in naturalistic settings. In [27], Pharo and Järvelin suggest systematic use of the triangulation of different data collection techniques as a general approach to get better knowledge of the Web information search process. However, the preparation of the setting for this kind of experimental study is not easy and only a few works explicitly claim the need of a combination of different sources of evidence. Examples of the complexity of these studies can be found in the following studies: [9,20,21]. Basically, they all suggest that further studies, including interview-based examinations of small groups of users, or laboratory tests or questionnaires to answer more complex questions should be carried out.

Studies that present an approach combining different sources of data include:

- Assadi et al. [10], which presents an approach which relies on a combination of three methods: user-centric data analysis, online questionnaires, and interviews.
- Teevan et al. [29], where the authors show how to combine two different streams of data, search query logs, and click-streams, to analyze re-finding behavior of a group of users under observation for a period of 1 year.
- Grimes et al. [15], where Google researchers present a work which shows that three different modes of collecting data—the field study, the instrumented user panel, and the raw query log—provide complementary sources of information. The main claim of the authors is that fully understanding user satisfaction and user intent require a depth of data unavailable in search query logs, but possible to acquire from other sources of data, such as one-on-one studies or instrumented panels.

2 Motivations

The comprehensive study we conducted was in the context of the TELplus project,[2] whose overall goal was to strengthen, expand, and improve TEL. Managed by CENL,[3] TEL is an on-line portal providing access to the electronic resources of the National Libraries of Europe. During the past years, TEL has gradually expanded to include 48 National Libraries from all EU[4] and EFTA[5] countries, this means that TEL offers access to the resources of 48 national libraries of Europe in 35 languages.

The resources accessible through TEL can be both bibliographical records and digital objects (e.g., books, posters, maps, sound recordings, and videos). The amount of referenced digital collections is constantly increasing. Quality and reliability are guaranteed by the professional skills of the librarians which operate in the 48 collaborating national libraries of Europe.

The different categories of users of the TEL portal can vary in terms of the type of services and documents they would like the system to provide. Therefore, a sound understanding of these needs and preferences was considered essential to ensure the quality, as well as the interoperability of the services provided. We considered that a detailed analysis of user interaction with TEL would provide interesting data about user preferences. This would then allow us to "learn" how to improve the prototype designed to better fulfill those preferences—e.g., by presenting information results in different ways, by allowing the user to configure the system according to their contextual needs.

This study reports on results that have been reached in the last part of a period of analysis of the TEL portal which lasted for more than 3 years. At the beginning of the collaboration within the TELplus project, a general methodology for gathering and mining information from Web log files was proposed. A series of tools to retrieve, store, and analyze the data extracted from log files were designed and implemented as reported in [5]. In [3], the authors report on initial findings on a specific aspect that is highly relevant for personalization services: the study of Web user sessions. Preliminary results about the added value gained by analyzing data derived from the chosen and combined sources of data were presented in [4]. A more structured analysis about the possible combination of implicitly and explicitly collected data to improve understanding of user behavior was presented in [6]. In [8], the initial study of a combined analysis of two different logs—action logs and HTTP logs—was presented. In addition, the first batch of user study was

[2] http://www.theeuropeanlibrary.org/telplus/.

[3] http://www.cenl.org/.

[4] http://europa.eu/.

[5] http://www.efta.int/.

conducted to collect into HTTP logs enough data to study the browsing activity and analyze possible relations between explicit preferences collected by online questionnaires and implicit actions recorded in the logs.

The user survey, which was initially designed and used to report preliminary findings [4], was re-designed using the lessons learned with its use in conducting the first user survey.

A revised version of the user survey was developed and used in a subsequent and complete study and the results of this comprehensive study are here reported.

The article is organized as follows: Sect. 3 explains the rationale for collecting the log data used in the TEL study. Section 4 presents the methodology and design of the user study conducted. Section 5 reports on the main findings and factors affecting the user interaction with the TEL portal. Section 6 derives useful indications from the study results that are of general use for the designers of Web portals. Section 7 provides some final remarks and possible generalizations of the method and analysis conducted. Section 8 concludes this article by mentioning future research initiatives related to this study's outcome. The last part of this article is the questionnaire designed and used to collect data on the use of the users that approach the European Library Web site/portal for the first time.

3 Use of log data

Log data are typically collected for understanding usage patterns of different software applications, with the aim of better adapting them to the needs, preferences, and expectations of their target users [22]. In the context of evaluating DL services they can provide useful data to implicitly observe usage behaviors over the long term and support the discovery of interesting phenomena regarding user interaction with the different types of services offered [1]. The query log is the less rich source of data for individual events, but it does provide essential information for understanding the scope of the resources that a search engine needs to provide to the user [15].

Even though explicit and implicit methods for investigating user needs and preferences are usually deployed in a separate way, different authors have recommended their combined use [4,17,27]. In particular, the combined sets provide the opportunity of reaching insights towards user personalization of Web searching services.

In the studies we conducted, the analyzed and combined sets of data provided the opportunity of reaching better insights not only into the needs and preferences of the participants initially involved, but also into other relevant communities of users potentially interested in using TEL services, as well as into their needs for more personalized offers of these services [7].

4 User study

Comprehensive survey studies conducted within long-term investigations of digital library services can support a deeper understanding of user needs, preferences, and behavior in different usage conditions, and can give directions to better design Web portals and the systems that are behind the portals. Since, surveys require a significant amount of time and effort to be carried out, a thorough design and plan for their deployment needs to be developed [19].

In the following sections, we describe the study context and the selected user group that was to be representative of the user community of the TEL portal, the questionnaire designed to interview users about their initial interaction with the TEL services, and the main task. The questionnaire is reported at the end of this article.

4.1 The study context and user group

The experimentation was carried out within a large-scale educational setting: the participants were requested to use TEL services in computer laboratories available at three different faculties of the University of Padua, Italy.

The participants were 216 students registered at the Bachelor and Master Programs of the three faculties involved, which were the Faculty of Humanities, Psychology, and Statistics. Most of the participants were of Italian nationality, equally distributed for gender and age range (between 19 and 25 years old).

The three faculties represent three different and major knowledge areas: humanities, social and medical science, and science.

The selected user group is representative of the users interested in the TEL services, because the contents made available through the TEL Web portal are those of European national libraries, which are natural reference libraries of university students.

The overall duration of the study lasted from October 2007 to June 2008. This is the annual time interval of lecturing of the University of Padua, where the lectures of the first semester begin in October and end in December or January, and the lectures of the second semester begin in March and end in June.

A detailed description of the class and type of undergraduate/Master courses attended by the study participants is presented in the following list:

– Basic course on foundations of computer science:

  AMS: a 3-year undergraduate course in Visual and Performing Arts;
  STB: a 3-year undergraduate course in History and Conservation of Artistic and Cultural Heritage;

Table 1 Number of students who participated in the study divided by course

Course	AMS	STB	LCM_MLC	ST_LE	Psychology	Statistics	LSLSC	Total
Students	23	16	48	65	22	6	36	216

LCM_MLC: a 3-year undergraduate course in Modern Languages and Literature and Linguistic and Cultural Mediation (class of "Informatica Generale");
ST_LE: a 3-year undergraduate course in History and Arts and Humanities (class of "Informatica generale");
Psychology: a 3-year undergraduate course in Cognitive Psychology and Psychobiology.

– Advanced course in database management:

  Statistics: a Master course in Statistics and Computing Technology.

– Introductory course in database management and Internet applications:

  LSLSC: a Master course in Foreign Languages for International Communication.

Table 1 reports the number of participants, divided by type of course attended.

4.2 The survey questionnaire

The questionnaire is composed of five main parts:

1. Identification of the target user and Web browser: to track the activity of each student recorded in TEL log files, it was necessary to change the string that identified the user-agent of the Web browser. The questionnaire reported clear instructions initially guiding the user to perform three easy actions to correctly set up the log system. The data recorded in the log files were also intended to support the subsequent design of profiling and personalizing features for the TEL portal.
2. Recording of the first action carried out: the questionnaire prompted students to report the first action carried out when they first accessed the TEL portal and its homepage was displayed. This question was intended to provide interesting feedback regarding user expectations and first reactions towards the TEL portal services and interface.
3. Browsing and search activities: nine multiple-choice questions (including nested sequences) were asked to uncover what types of searches participants made, which areas of the Web portal they had browsed, and what was their perceived usefulness of the sections visited.
4. Understanding of the Web portal and reasons for disappointment: two questions more specifically addressed user perception of the clarity of the information services offered by the Web portal and any possible causes of disappointment with it.
5. Registration with the portal and Web browser reset: the last four instructions reported in the questionnaire were intended to guide users to complete their registration with the Web portal for future analysis, to reset the original user agent of the Web browser and delete personal data.

The questionnaire consisted of 17 main actions and questions, and was provided in PDF format[6] with a direct link for email submission (see questionnaire form in Appendix). The data collected were automatically transformed into XML format[7] for subsequent analysis.

All participants in this study had never used the TEL portal before. During the questionnaire administration a facilitator from the University of Padua was present with the aim of:

– providing a brief introduction and motivation to the study;
– providing technical support, if required;
– taking notes regarding participant questions, comments or any critical event that might have been useful to consider during the following data analysis.

The questionnaire administration was conducted during special laboratory hours alongside of course lectures. This means that the sessions were conducted in different days over the time span October 2007–June 2008.

4.3 The task

Participants were assigned the task of performing a free navigation and search for information on TEL portal, and then to fill in the electronic questionnaire assessing their level of satisfaction with the different sections of the portal experienced. The only help the students got in case something went wrong was during the set-up of the user agent string of the browser necessary to identify each single user (steps 1, 16, and 17 of the questionnaire).

[6] http://www.adobe.com/devnet/pdf/pdf_reference.html.

[7] http://www.w3.org/XML/.

Table 2 Summary of statistics for the time of a user session in minutes calculated using the HTTP logs, and the search logs

Type of log data	HTTP log	Search log
Median	1.3	2.0

The choice of not assigning predefined navigation and search tasks to users was due to our principal interest in understanding users level of motivation and expectations when first interacting with the portal, without affecting their initial willingness to explore any section of the portal (including the help section) to get an idea of the type and quality of services offered. This would also make our study conditions and observations more comparable to the results of the log data analysis conducted in [2], helping us to shed light on the reasons for users typical low engagement with the system, as observed before.

5 Analysis of results

The analysis of the data gathered from the questionnaire focused on relevant aspects of the user interaction that were derived from the log data. In particular, we were interested in understanding the possible causes that lead to very short sessions both for the browsing activity and the search activity. The different activities that are carried out during browsing were analyzed using the records of the HTTP log data. The search activities were analyzed making use of the search log data that contains relevant information on user queries. The median length of sessions per user is reported in Table 2. This length suggests that users may have encountered some problems during the interaction with the portal, e.g., difficulties in learning how to read the list of the results provided by TEL to their queries, as well as some problems in learning how to properly manage the interface functionalities.

Therefore, the analysis of the questions was conducted after clustering the most relevant questions in the following four groups:

1. Action performed when entering the portal: the results were mainly derived from the analysis of answers to question 4;
2. Language issues: mainly derived from the analysis of answers to question 5;
3. Multimedia documents: mainly derived from the analysis of answers to question 7;
4. Disappointment or satisfaction about the offered services and results: derived from the analysis of answers to questions 6, 12, and 13.

This same grouping of topics will be used in the following sections to present the results of the data analysis conducted.

5.1 Action performed when entering the portal

Question 4 asks the user to record the first action performed when entering the portal. The Pareto diagram for the answers is shown in Fig. 1a (the seven students who did not answer this question are not counted). A Pareto diagram is a bar chart of frequencies sorted by frequency and it is used widely in quality control settings to identify critical factors leading to failure or defects in a process [31]. In our case, this diagram highlights the trend of the frequencies in the answers and emphasizes the differences between choices.

About 56% of people changed the language of the interface to make the browsing and understanding of the Web site easier. The other four choices all together have more or less the same weight.

From these numbers it is possible to make some considerations if we take into account the fact that the users of this study are students with a limited knowledge of the English language. More than 50% of the people who connect to the portal decided to switch the language of the interface from default (English) to their mother tongue (Italian), confirming the natural user tendency to recreate a more familiar environment when facing the use of a new system. This is even more evident in settings where novice users have to deal with technical terms or neologisms in another language.

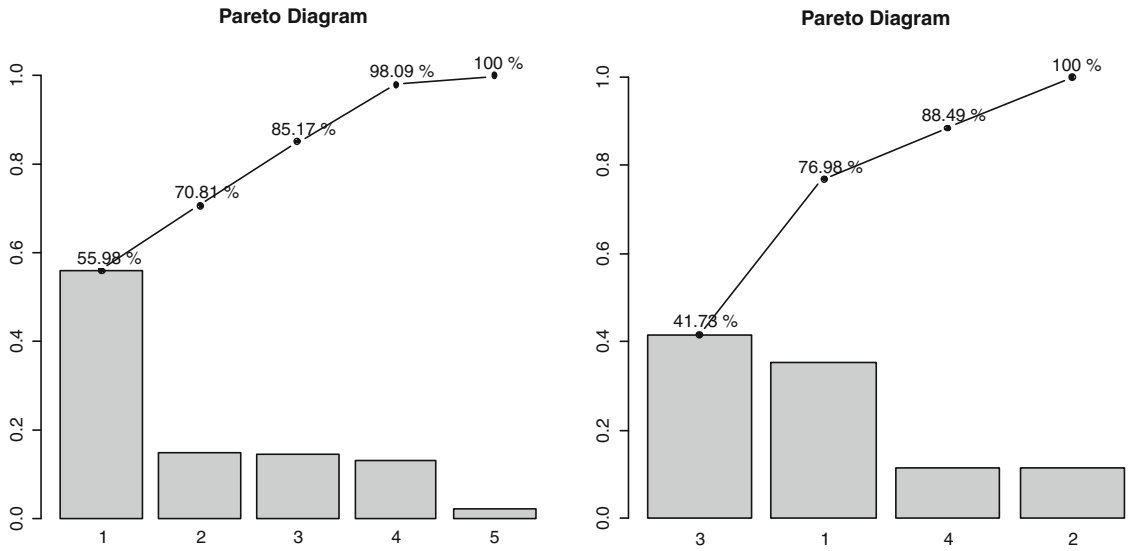
This result is of interest to all designers of Web portals and services that can be offered to users that use different mother tongues.

5.2 Language issues and quality of results

Question 5 asks whether the switch of the interface language made the interaction with the portal easier.

Although, the majority of people answered “Enough” (40.21%), we wanted to investigate whether the total of the answers represented a mild or strong indication of a positive or negative feeling. For this reason, we performed a statistical test to verify the hypothesis that, on average, the answers are equal to the central value “Enough”, corresponding to point 3 of a Likert scale (1–5). The result of the t test showed that answers are not significantly different from the central value, which confirms that there is no agreement on the opinion users expressed on this question. This result can be interpreted in this way: the switch of the interface language, that is in general the first thing a user does when accessing the portal, does not clearly help users in the navigation of the portal.

The fact that switching the language does not help the user may be explained by the poor quality of the translation provided by the portal. In fact, it has been noticed that for some languages the quality of the translation was not as good as English (the default language) or in some cases the translation was completely missing, a situation which leaves the user quite disoriented and disappointed.



(a) Question 4. Pareto Diagram which Reports on Action Performed when Entering the Portal.

(b) Question 13. Pareto Diagram which Represents Results on Possible Causes of Disappointment.

Fig. 1 Derived Pareto diagrams for questions 4 and 13. On the x-axis the value of the Likert scale of each question, on the y-axis the percentage of answers. **a** Question 4. Pareto diagram which reports on action performed when entering the portal. **b** Question 13. Pareto diagram which represents results on possible causes of disappointment

Question 6 addressed the search activity. We focused the analysis on the answers to the subquestion "Are the results satisfying?".

The answers gave important clues about one of the two most disappointing things about the TEL portal: the results or their contents. Around 60% of the students considered the results in a neutral/moderately negative way, and around 80% of them expressed a moderate negative consideration of the results retrieved during the search. Such a large group of users disappointed by the results should be considered carefully, and for this reason we performed a t test to verify the hypothesis that on average the answers are equal to the central value "Enough". The result of the test shows that the probability that the answers are not different from the central value is so small that we have to reject this hypotheses; therefore, answers are significantly different from the value "Enough" and on the negative side, since the average value is 2.54.

5.3 Multimedia documents

Question 7 is composed by a sequence of subquestions. We focused on the first two subquestions: "Is the object in the image clear enough?" and "Did you find the treasures section interesting?" because they are the two most relevant ones in relation to the "Multimedia" analysis. These two subquestions are strictly correlated and refer to one of the sections of the portal that were visited most frequently by users.

For the first subquestion, more than 80% of the users answered positively ("Very" or "Enough"), and this fact is confirmed also by the statistical test carried out on the data: the probability that the answers are not different to the central value 3 ("Enough") is almost zero; therefore, we can conclude that users clearly stand positively.

For the second subquestion, about 71% of students were moderately happy about the "treasures" section. Again, to understand better whether there is a different distribution of satisfaction among students, a t test was carried out with the hypothesis that there is no significant difference between the answers and the central value 3. The p value of the test is very close to zero, which means that we have to reject the hypothesis and conclude that on average there is a clear intention of the users to judge positively the section which shows the treasures.

It is important to underline that around 25% of the students did not answer both questions. This may be interpreted as showing the willingness to conclude the questionnaire quickly or their lack of interest in exploring the portal more deeply. For this reason, we cross-analyzed these results with the answers given to the first two subquestions of question 6 which investigate the type of search and query. The result is that 23 subjects declared they did not activate any search nor examined the treasure section; we may consider these subjects not interested in the contents provided by the portal. Another group of 33 subjects chose not to search or issue any query, but to visit the "Treasures" section. However, about 80.8% of users who answered both questions did both a search and made a visit to the treasures section.

5.4 Disappointment or satisfaction about the offered services and results

5.4.1 Clarity of the portal and causes of disappointment

Question 12 intends to capture what is the first impression of the user about the portal, and Question 13 the possible causes of disappointment.

A *t* test on the available data on the answers to Question 12 was performed to verify the hypothesis that the average answer is equal to the central value ("Enough"). The result of the test is that we cannot reject the hypothesis with a very high *p* value. This means that there is no clear and definite position taken by the students, and therefore it is only possible to assert that answers are distributed evenly around value 3. The distribution of the answers, which is perfectly balanced between the positive and negative answers shown in the first column (Question 12) of Table 3, suggested a further analysis which consisted of correlating the answers of this question to Question 13 about the possible reasons for disappointment. Before we present the analysis of the two questions together we need to present the results of the analyses of Question 13.

The Pareto diagram for the answers to Question 13 (excluding the 77 students who did not answer) is shown in Fig. 1b which shows that about 42% of the students who answered this question said that there was something disappointing in the results retrieved. Immediately after, there is another large group of students who were disappointed by the portal interface. Therefore, the majority of people, more than 75%, are unsatisfied by either the interface or the contents.

However, the number of users who did not answer Question 13 is not negligible: 77 out of 216 subjects, which means that more than 35% of users decided not to answer. This number is so unexpected that we decided to investigate this matter more deeply, to understand whether it was possible from the available data to discover hidden meanings: Was a missing answer the indication that overall there was no particular reason for being unsatisfied? Are there any reasons that can be brought to light? To analyze these facts, we compared answers to the first subquestion of question 6 ("What were you looking for?"), the second subquestion of question 6 ("What was the query?"), and the given alternative answers to the question which investigates the cause of disappointment, and we found that 30 subjects out of 77, equal to 39% of people, who did not answer the first alternative of Question 13, did not even answer the previous subquestions. This agreement in not answering all these three subquestions makes the following interpretation plausible: those 39% of people who did not answer the first alternative of Question 13 and declared that they did not perform any search or query are likely to be a type of "random-user" who examine the content of the portal without having any task or objective to accomplish. The case is different for those 47 out of 77 people who did not answer the first alternative of question 13, but performed some search: this time the fact that they did not answer may indicate the "absence of reasons for being unsatisfied".

Going back to the 30 users who did not perform any search, besides not answering the possible cause of disappointment, we also analyzed the kind of actions they declared to perform as first action (first alternative of Question 4). Results indicate that these users:

- modified the interface language: 17 (58.63%)
- visited the Treasures section: 7 (24.14%)
- performed other actions (not search): 5 (17.24%).

Coherently, none of the subjects indicated any kind of action related to search.

To analyze the answers of both Questions 12 and 13, we divided students into three groups: see Table 3 for users who did not understand completely what the portal offers ("Not at all" and "A little"), users who did not take a clear position ("Enough"), and users who understood clearly

Table 3 Breakdown of answers to questions 12 and 13

Question 12			Question 13				
			Web Interface	Content	Search results	Other	Not answered
"Not at all"	6	27.31%	19	10	16	3	11
"A little"	53						
"Enough"	89	41.21%	24	4	25	7	29
"Very"	52	27.31%	5	2	16	6	30
"Highly"	7						
Not answered	9	4.17%	1	0	1	0	7
Total	216		49	16	58	23	70
			22.68%	7.41%	26.85%	10.65%	32.41%

what the portal offers ("Very" and "High"). For each group, we studied the answers to Question 13 and found possible explanations.

For the 59 participants who did not understand what the portal offers ("Not at all" and "A little"), the two main causes of disappointment were the interface and the search result, 19 and 16 students, respectively. It is interesting to see how important the interface for the overall judgement of the portal is, especially if it is cross-analyzed with the problem of the interface language. The content provided by the portal is another main cause with 10 users, which is the highest relative number compared to the other two groups of students. Finally, it is not easy to interpret the reactions of users who did not answer question 13: these are students who declared they did not understand the portal, although they did not provide any reason for that.

Many of the other 59 students who did understand the portal ("Very" and "High") decided to answer the first alternative of question 13. This is a valuable set of data since it represents those students who were satisfied by the portal and wanted to give indications for a possible improvement of the Web site: 16 students declared that the main cause of dissatisfaction was the search results, 13 students indicated different aspects of the interface, contents and others. The 30 students who did not answer the alternatives can be interpreted as users who liked the portal and did not find anything particularly wrong with it.

There is a relative majority of students undecided about what the portal offers ("Enough") who gave clear indications about the possible features to improve. Regarding the alternatives of question 13, answers were once again evenly distributed among the interface and the search results, while a large portion of people did not answer. The fact that this group of users did not answer is even more difficult to interpret compared to the other two groups, since it is not clear whether this choice indicates that in general their user experience with the portal was good, bad, or maybe not worth commenting.

The number of users who did not answer this question was small: 9 participants (about 4% of the total sample).

5.4.2 Further considerations about causes of disappointment

One main issue raised was related to the languages option. It is preferable to interact with the portal in your own language instead of the default English language. However, this does not exclude the fact that users search and browse collections of documents written in different languages, as it was also observed by the study facilitator that attended the students sessions. This behavior may change over time, especially with the constant usage of the portal. Images and multimedia documents in general seem to be very appealing for users; the "Treasures" section, which presents high resolution images of ancient documents, and the "Exhibitions" section, which provides pictures of the national libraries buildings, were very often browsed by users even before trying any query to the portal. For this reason, providing a direct linking from the images to the catalogue records is recommended, since currently there is no possibility of browsing from an image to the collection in which it is stored.

The most common answers to questions 12 and 13 may help us to interpret other main causes of dissatisfaction: the interface and the quality of the search results are not satisfactory for more than half of the users who do not understand or are uncertain about what the portal offers. The problem with the interface is less serious for users who liked the portal, but most of these users declared they had problems with the search results.

Comments collected from some participants allowed us to understand that for a generic user who has never seen the TEL portal, it is not so clear what kind of information can be accessed. Retrieving library catalogues after a search, instead of Web pages or digital documents, is disappointing and motivates the user to leave the portal quickly.

The interface is also important: first, not every section or subsection is translated entirely; second, the search part is not clear for users who are used to interacting with very simple and basic Web search engine interfaces (e.g., Google). The interaction with the portal and the clarity of what is offered is still an issue for TEL.

The behavior of a large part of students—32.41% out of the total—who did not give any indication about their possible causes of disappointment is more difficult to interpret. This result could be interpreted either as "nothing wrong to point out" by students who liked the portal, or as a lack of clear opinion and understanding about the portal services after the first interaction.

A main reason for the short duration of user sessions is the general disorientation felt by the user upon entering TEL portal for the first time. We speculate that for users already familiar with web search portals (e.g., Google Scholar) where it is possible to search and retrieve the documents you are looking for (e.g., links to relevant web pages or online documents), the possibility of accessing only lists of bibliographic records through TEL could be a bit disappointing, and in some cases lead to leaving the portal quickly.

The presentation of query results in a more visual format might be more appealing and stimulating for novice users who access the system. Our log data show that both the "Treasures" section and the "Exhibitions" section were thoroughly browsed by users even before submitting any specific query to the portal. This would suggest as a possible TEL improvement to provide more linking from catalogue records to images, when this is possible.

We are aware that our sample of participants to the survey study cannot be representative of all potential categories of TEL users, since each community brings different kinds of expertise, preferences, and work practices. Nevertheless, the findings collected help us to understand how to improve the TEL services and interface to fulfill the needs and preferences of at least one important target user group within the academic environment. These findings provide a good starting point for deciding how TEL services can be better customized to overcome the difficulties met by the survey participants and how to effectively monitor the evolution of TEL usage patterns over time.

6 Lessons learnt for a better design of Web portals

The methods used to conduct the study and the results of the study itself give useful indications that can be of general use for the designers of Web portals.

6.1 User habits influence

The study on novice user interaction with the TEL portal has confirmed previous findings regarding the importance of considering web users searching-navigation habits when providing new digital library services and tools. In [30], undergraduate students of the University of Linz (Austria) were found to prefer Google search to Prospector, a personalized meta-search engine, since the new system did not offer results in German (the mother tongue of all the study participants). In the case of our participants, it was important for them to change the interface language to their mother tongue, and to directly retrieve the documents or information looked for instead of their bibliographic records (as enabled by Google search).

The influence of previous user habits is particularly strong when the user has a limited knowledge of the relevant system properties, e.g., the kind of services offered by a digital library portal. Although, we can expect that this knowledge might increase over time, it is important not to disappoint users during their first interaction with a new system, since this might negatively affect their intention to use the digital library portal in the future.

6.2 Improving interface clarity and appeal

Improving the interface clarity and appeal would be a good strategy to attract users, stimulating their acquisition of familiarity with the new system by exploiting the so-called "novelty" effect, thus encouraging further exploration. The system interface should make particularly visible and easy to understand the different choices available, among those the possibility of customizing the new system to set up a more familiar environment, and the type of services made accessible through the new system.

Observations of users first interaction with the TEL portal showed that visual and multimedia documents in general were found particularly appealing. As previously underlined, the "Treasures" section as well as the "Exhibitions" section were very often browsed by users even before trying any other query to the portal. For this reason, a required improvement would be the linking from visual documents (e.g., images) to the catalog records, since the current TEL system version does not support browsing from an image to the collection in which it is stored.

6.3 Web portal motivations must be explicit

When it comes to understanding the possible causes of use or not use of a newly developed system, differences among user preferences and attitudes towards relevant technology should also be borne in mind. As reported in [28], different varieties of non-use, beyond use, can also be identified and designers of new digital library services should get a more detailed picture of their underlying characteristics and motivations.

The survey results show that some participants might develop a more active resistance towards the TEL portal, that is, they might refuse to adopt it in a more active and intentional way.

A possible reason can be their preference for alternative modalities of interaction and engagement with digital library services, for example by means of more visual and interactive interfaces they are already familiar with. It is likely to expect that these groups of non-users will be more active than others in finding alternative ways of accessing the information technologies and services required, as well as in providing comments and information regarding their reasons for not appreciating the new system provided, e.g., by expressing strong positions in the satisfied/disappointed rating scales of the questionnaire, recommendations on how to improve the system and so on.

Other participants (e.g., 32% of our sample) show a general disinterest towards the new system, little exploration of the features and options available, together with missing answers to parts of the post-session questionnaire. Getting information on how to improve the tool and services offered to meet their requirements is a more challenging task.

7 Generalization of the study methods and results

The case study presented is part of our current effort towards developing a more comprehensive and long-term framework for the requirements collection, analysis, and evaluation of digital library Web portal and services.

As we have shown, implicit data on user interaction with a new DL portal were useful for detecting typical usage patterns of different user groups, as well as their possible evolution or stability over time. They provided key information for identifying possible interaction problems on a large-scale (e.g., the whole user population), but little or no insight as to why these problems occurred and how they could be solved. They undoubtedly were influential in the design of more focused survey studies (based on explicit user data) able to investigate which relevant user expectations, habits, motivations, preferences, or difficulties were responsible for the interaction patterns observed.

Although, the design of effective survey tools and studies for collecting user data is always a complex task, we believe that combining both implicit and explicit data collection methods, together with establishing an effective information flow between these two types of information sources in a longitudinal perspective, can greatly improve the quality of the evaluation work conducted.

From a more comprehensive understanding of the reasons for typical user interaction with DL services and interfaces, it is then easier to cope with the system refinement phase by deciding which improvements have to be introduced first to fulfill the needs and preferences of our target users. The data collected allow us to more easily interpret, compare, and possibly extend the survey findings and approach developed to other types of DL systems and users in the near future, and thus to further proceed with our longitudinal investigation.

The case study described has helped us to achieve a more complete understanding of the user experience with a new DL Web portal and services, as well as raise our awareness about possible improvements to the data collection methods deployed, which will feed our next evaluation efforts.

8 Conclusion and future study

In this article, we presented a study of complementary types of logs in complex systems with the aim of finding new ways of using them to evaluate and personalize digital library services for the final users. Studies on log files are essential for personalization purposes, since they implicitly capture user intentions and preferences in a particular instant of time. Computer science researchers have been building a case for search log access to allow them to study and analyze new information retrieval algorithms via a common benchmark search log, as well as to learn about user information needs and query formulation approaches. Social scientists could investigate the use of language in queries as well as discrepancies between user interests as revealed by their queries versus user interests as revealed by face-to-face surveys.

This study also gave us the opportunity of discussing the availability and use of log data: considerations included how log files should be made publicly available to researchers, whether log data should be gathered for specific tasks, whether there is value in general log data, how additional information can be gathered and correlated with query log data. The current lack of recent and long-term query logs makes the verifiability and repeatability of log analysis experiments very limited. A first attempt in this direction has been made within the Cross-Language Evaluation Forum in 2009 and 2010, the effort is named LogCLEF and it is aimed at stimulating research on user behavior in multilingual environments and promoting standard evaluation collections of log data. In contrast to traditional IR benchmarking activities, LogCLEF leaves room for individual approaches to log analysis [13].

The experience that has been gained with this study is proving to be sufficiently general to be applied in other contexts e.g., the CULTURA project[8] where the characterization of users and communities that interact with the CULTURA environment is under way and is making use of the presented findings here.

Acknowledgements The authors would like to thank the staff of TEL Office for their systematic support during the analysis of logs which is reported in the article. The authors are most grateful to the anonymous associate editor and the referees for their very helpful comments. The authors like to thank Alexander Cormack, whose revisions have added to the quality of the study presented here. The study has been partially supported by TELplus Targeted Project for digital libraries, as part of the eContent *plus* Program of the European Commission (Contract ECP-2006-DILI-510003). The study is also partially supported by EuropeanaConnect (Contract ECP-2008-DILI-52800).

Appendix

[8] http://www.cultura-strep.eu/.

Submit by Email

Name_____ Surname_____ Student ID_____ Computer_____

TELplus: First Use of The European Library Web Site

Author: UNIPD
Version: 2.0

1. Use the following instructions to change the identification of the Web browser:

- Type *about:config* as address in the address bar of Mozilla Firefox, the location where you normally enter a URL and press *Enter*

- Right click to open the context menu and select "String" from the menu entry "New"

- Type the preference name *general.useragent.override*

- Type the new User Agent value you want Mozilla Firefox to use, and insert "faculty_date_computer-number" (e.g. lettere_20080101_pc03)

- Check the new value by typing *about:* in the address bar

These operations enable the "User Agent" to be modified and allow the activities performed to be identified.

The User Agent is a text string which represents and identifies who is interacting with the Web Server. The User Agent is usually composed by the name and the version of the browser.

2. **Date** _____ **Starting Time** _____

3. You are now invited to spend some time to familiarise yourself with the service offered by *The European Library* Web site, which is a free service that offers access to the resources of 47 European national libraries in 20 different languages. Resources can be both digital resources and bibliographical records (books, posters, maps, sound recordings, videos, etc.). The URL to type is:

http://www.theeuropeanlibrary.org/

4. What was the first thing you did when you entered *The European Library* Web site?

◯ Change the language of the interface

◯ Type a query in the search box

◯ Click on the "treasures"

◯ Choose another section. If so, specify which one: _____

◯ Other _____

5. If you modified the language of the interface, did that make access to the Web site easier?

1	2	3	4	5
◯ Not at all	◯ A little	◯ Enough	◯ Very	◯ Highly

6. If you typed a query in the search box:

What were you looking for?

_____

What was the query?

_____

Did you use any search criteria?

_____

Was the visualization of the results clear?

1 2 3 4 5

◯ Not at all ◯ A little ◯ Enough ◯ Very ◯ Highly

Did you browse the results? ☐

What did you expect from this search?

_____

Were the results satisfying?

1 2 3 4 5

◯ Not at all ◯ A little ◯ Enough ◯ Very ◯ Highly

Did you consider the effects of the "stop the search" function while waiting for the results of a query? ☐

7. If you selected the "treasures":

Was the object in the image clear enough?

1 2 3 4 5

◯ Not at all ◯ A little ◯ Enough ◯ Very ◯ Highly

Did you find the "treasures" section interesting?

1 2 3 4 5

◯ Not at all ◯ A little ◯ Enough ◯ Very ◯ Highly

Did you enlarge the image? ☐

Has the item “View treasure" been consulted? ☐

Other _____

8. If you selected another section, what did you elect in particular?

_____

Did you find the section interesting?

1 2 3 4 5

◯ Not at all ◯ A little ◯ Enough ◯ Very ◯ Highly

9. In case you selected the section *collections,* was it clear how the section is organized?

1 2 3 4 5

◯ Not at all ◯ A little ◯ Enough ◯ Very ◯ Highly

What collection did you choose?

_____

Did you find the collection section interesting?

1 2 3 4 5

◯ Not at all ◯ A little ◯ Enough ◯ Very ◯ Highly

10. Did you use the "help?" section?

Did you find the help section interesting?

1	2	3	4	5
◯ Not at all	◯ A little	◯ Enough	◯ Very	◯ Highly

11. Did you use the "Sitemap"?

Did you find the Sitemap interesting?

1	2	3	4	5
◯ Not at all	◯ A little	◯ Enough	◯ Very	◯ Highly

12. Upon a first impression did you understand what the Web site offers?

1	2	3	4	5
◯ Not at all	◯ A little	◯ Enough	◯ Very	◯ Highly

13. If you feel disappointed, do you think the cause is:

◯ The interface of the Web site

◯ The contents

◯ The search results

◯ Other _____

14. Use the following instructions to become a registered user:

- Return to the home page of the Web site

- Click on the Register button

- Follow the instruction displayed

- Remember the username and the password you choose

15. Ending time

16. Use the following instructions to reset the user agent:

- Type *about:config* as address in the address bar and press *Enter*

- Type *general.useragent.override* in the space "Filter".

- Select the line and right click.

- Select "Reset".

- Type in the address bar "about:" to check the default value of the user agent.

17. Use the following instructions to reset the browser and delete personal data:

- Press together Ctrl+Shift+Canc on the keyboard.

- Select only "cache" and "cookie".

- Click on "Delete personal data".

References

1. Agosti, M.: Log data in digital libraries. In: Agosti, M., Esposito, F., Thanos, C. (eds.) Post-proc. of the 4th Italian Research Conf. on Digital Library Systems, IRCDL 2008, Padua, pp. 115–122. DELOS, Delos (2008)
2. Agosti, M., Di Nunzio, G.M.: Gathering and mining information from web log files. In: Thanos, C., Borri, F., Candela, L. (eds.) DELOS Conference, Lecture Notes in Computer Science, vol. 4877, pp. 104–113. Springer, Berlin (2007)
3. Agosti, M., Di Nunzio, G.M.: Web log mining: a study of user sessions. In: 10th DELOS Thematic Workshop on Personalized Access, Profile Management, and Context Awareness in Digital Libraries. http://www.dblab.ntua.gr/persdl2007/papers/72.pdf (2007). Last accessed 27 Sept 2011
4. Agosti, M., Angelaki, G., Coppotelli, T., Di Nunzio, G.M.: Analysing HTTP logs of a European DL initiative to maximize usage and usability. In: Goh, D.H.L., Cao, T.H., Sølvberg, I., Rasmussen, E.M. (eds.) 10th International Conference on Asian Digital Libraries, ICADL 2007, Lecture Notes in Computer Science, vol. 4822, pp. 35–44. Springer, Berlin (2007)
5. Agosti, M., Di Nunzio, G.M., Niero, A.: From Web log analysis to Web user profiling. In: Thanos, C., Borri, F. (eds.) DELOS Conference 2007. Working Notes, pp. 121–132. Pisa, Italy (2007)
6. Agosti, M., Crivellari, F., Di Nunzio, G.M.: Digital library services using complementary logs. In: Workshop on Understanding the User, Logging and Interpreting User Interactions Search and Retrieval. Held in Conjunction with the 32nd ACM SIGIR Conference, Boston, MA (2009)
7. Agosti, M., Crivellari, F., Di Nunzio, G.M.: A method for combining and analyzing implicit interaction data and explicit preferences of users. In: Workshop on Contextual Information Access, Seeking and Retrieval Evaluation. Held in Conjunction with the 31st European Conference on Information Retrieval (ECIR 2009), Toulouse (2009)
8. Agosti, M., Crivellari, F., Di Nunzio, G.M., Ioannidis, Y., Stamatogiannakis, E., Triantafyllidi, M.L., Vayanou, M.: Searching and browsing digital library catalogues: a combined log analysis for the European library. In: Post-proceedings of the Fifth Italian Research Conference on Digital Library Systems, IRCDL 2009, Padua, 29–30 Jan 2008. DELOS: an Association for Digital Libraries, Delos (2009)
9. Anick, P.G.: Using terminological feedback for web search refinement: a log-based study. In: Proc. of the 26th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval, 28 July–1 Aug 2003, Toronto, pp. 88–95. ACM, New York (2003)
10. Assadi, H., Beauvisage, T., Lupovici, C., Cloarec, T.: Users and uses of online digital libraries in France. In: Koch, T., Sølvberg, I. (eds.) Proc. of the 7th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2003), Lecture Notes in Computer Science, vol. 2769, pp. 1–12. Springer, Berlin (2003)
11. Bollen, J., Luce, R.: Evaluation of digital library impact and user communities by analysis of usage patterns. D-Lib. Mag. **8**(6) (2002)
12. Borgman, C.L.: From Gutenberg to the Global Information Infrastructure. MIT Press, Cambridge (2000)
13. Di Nunzio, G.M., Leveling, J., Mandl, T.: Multilingual log analysis: LogCLEF. In: Proceedings of the 33rd European Conference on Advances in Information Retrieval, ECIR 2011, pp. 675–678. Springer, Berlin (2011)
14. Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., Klas, C.P., Kovács, L., Landoni, M., Micsik, A., Papatheodorou, C., Peters, C., Sølvberg, S.: Evaluation of digital libraries. Int. J. Digit. Libr. **8**(1), 21–38 (2007)
15. Grimes, C., Tang, D., Russell, D.M.: Query logs alone are not enough. In: Amitay, E., Teevan, C.G.M.J. (eds.) Query Log Analysis: Social and Technological Challenges. A Workshop at the 16th International World Wide Web Conference (WWW 2007), Banff (2007)
16. Hill, L., Dolin, R., Frew, J., Kemp, R.B., Larsgaard, M., Montello, D.R., Rae, M.A., Simpson, J.: User evaluation: summary of the methodologies and results for the Alexandria digital library. In: Proceedings of the ASIS Annual Meeting, vol. 34, pp. 225–243. University of California at Santa Barbara, Santa Barbara (1997)
17. Ingwersen, P., Järvelin, K.: The Turn. Springer, Dordrecht (2005)
18. Jameson, A., Gabrielli, S., Kristensson, P., Reinecke, K., Cena, F., Gena, C., Vernero, F.: How can we support users' preferential choice? CHI Extended Abstracts, pp. 409–418 (2011)
19. Kelly, D., Harper, D.J., Landau, B.: Questionnaire mode effects in interactive information retrieval experiments. Inf. Process. Manag. **44**(1), 122–141 (2008)
20. Koch, T., Ardö, A., Golub, K.: Browsing and searching behavior in the Renardus web service: a study based on log analysis. In: Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '04, pp. 378–378. ACM, New York (2004). doi:10.1145/996350.996444
21. Mahoui, M., Cunningham, S.J.: A comparative transaction log analysis of two computing collections. In: Borbinha, J.L., Baker, T. (eds.) ECDL, Lecture Notes in Computer Science, vol. 1923, pp. 418–423. Springer, Berlin (2000)
22. Nicholas, D., Huntington, P., Watkinson, A.: Scholarly journal usage: the results of deep log analysis. J. Documentation **61**(2), 248–280 (2005)
23. Nielsen, J.: Usability Engineering. Academic Press, Boston (1993)
24. Norlin, E.: Reference evaluation: a three-step approach-surveys, unobtrusive observations, and focus groups. Coll. Res. Libr. **61**(6), 546–553 (2000)
25. Norlin, E., Winters, C.: Usability testing for library websites: a hands-on guide. American Library Association, Chicago (2002)
26. Papatheodorou, C., Tsakonas, G. (eds.): Evaluation of digital libraries: an insight into useful applications and methods. Chandos Publishing, Oxford (2009)
27. Pharo, N., Järvelin, K.: The SST method: a tool for analysing Web information search processes. Inf. Process. Manag. **40**(4), 633–654 (2004)
28. Satchell, C., Dourish, P.: Beyond the user: use and non-use in HCI. In: Foth, M., Kjeldskov, J., Paay, J. (eds.) OZCHI, ACM International Conference Proceeding Series, vol. 411, pp. 9–16. ACM, New York (2009)
29. Teevan, J., Adar, E., Jones, R., Potts, M.A.S.: Information re-retrieval: repeat queries in yahoo's logs. In: Kraaij, W., de Vries, A.P., Clarke, C.L.A., Fuhr, N., Kando, N. (eds.) SIGIR, pp. 151–158. ACM, New York (2007)
30. Van Velsen, L., König, F., Paramythis, A.: Assessing the effectiveness and usability of personalized internet search through a longitudinal evaluation. In: UCDEAS-2009: 6th Workshop on User-Centred Design and Evaluation of Adaptive Systems, Trento. http://doc.utwente.nl/68858/ (2009). Last accessed 27 Sept 2011
31. Wilkinson, L.: Revising the Pareto chart. Am. Stat. **60**(4), 332–334 (2006). doi:10.1198/000313006X152243