Will Mobile Web Era impact the user's behaviour in a digital library?

Marco Scarnò[1], Donatella Sforzini[1], Ugo Contino[1], Paola Gargiulo[1]

[1]Cineca, Department of Information and Knowledge Management Services, Rome, Italy

Abstract: Almost 10 millions of articles were disseminated trough the web by CINECA.These are accessed by many academic Italian institutions located in the Centre-South of Italy. In order to be prepared for the "Mobile Web Era", it was realized an optimized version of the website to facilitate the "experience of the mobile users".
We analysed the behaviour of such users and of the standard ones; to do this we considered their "traces" stored into the web server log file. Some results are expected, like more browsing and less downloads; other surprising, like people that access the digital library only through a mobile device.

1. Introduction

There are many different definitions in literature on what digital libraries are; Lesk (1997) views them as "organized collections of digital information"; Francisco-Revilla et al. (2001) report digital libraries are increasingly being defined as collection pointers to web-based resources. We will refer to what suggested by Borgman (2002): these may be seen as new forms of information institutions, multimedia information retrieval systems that support the creation, the use and the searching of digital contents.

CINECA (formerly CASPUR) manages a digital libraries that is accessed from many Italian academic institutions; it contains more than 5700 academic and scientific full-text periodicals; these correspond to over 9.8 million of articles fully downloadable. Journals are available dating the nineties; they cover all fields and are issued by different publishers and professional societies, including, for example, the American Chemical Society, Blackwell Publishing, Elsevier Science, Institute of Physics Publishing, Kluwer Academic Publisher, Springer.

This service is accessible from a web site (http://periodici.caspur.it) that facilitates both the browsing of the content and the deep search of specific articles by means of a powerful engine that permits queries based on the author's name, on the titles, etc. This access is restricted to authorized Institutes and Universities through a procedure that checks the IP address or that considers a username and a password

By considering a digital library as a product that can be used by specific users to achieve specified goals, we would like to study the usability of our system from various perspectives.

Jeng (2005) states that usability can be grouped into two large categories: inherent usability and apparent usability. The first one in mainly related to the functional or dynamic part of the interface usability, by including those attributes that focus on how to make the product easy to understand, easy to learn, efficient to use, less erroneous and pleasurable. On the other hand, apparent usability is more related to the visual impression of the interfaces. It has to be noticed that both the concepts can be contradictory; for example, in a web page design, graphics enhance apparent usability but slow down the system.

The interest in usability of web sites is growing because of new devices that permit to access the Internet; according to Nielsen (2008) the penetration of mobile ones in Italy is 11.9 percent.

By taking into account this observation, CINECA,from June 2010, created a simplified access to the virtual library; this correspond to a new web site (http://m-periodici.caspur.it) that was conceived to facilitate its use within the most common mobile devices (like smartphones, tablet, etc.) or for their operative systems (IOS, Android, Windows Mobile). Fig. 1 shows the two different layoutof the web sites that permit to enter the CINECA digital library.
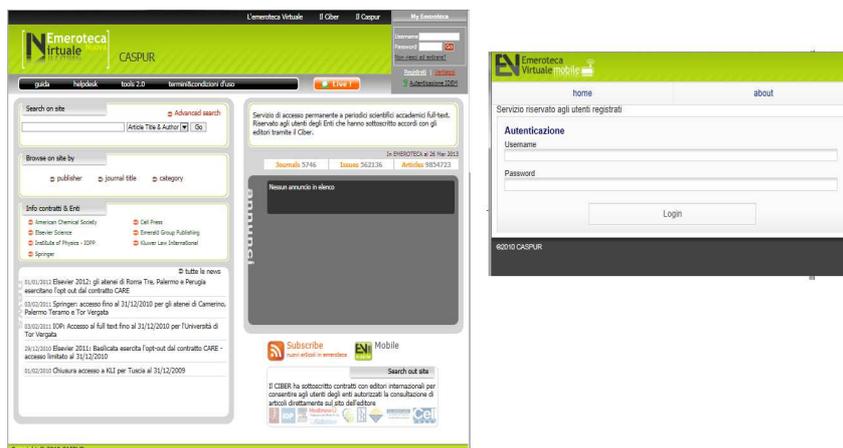

Fig. 1:the "standard" access to the CINECA digital library (on the left) and the one designed for the mobile users (on the right)

It has to be noticed that usability is often studied with a "qualitative approach", i.e. with questionnaires that investigate the users opinions, preferences, etc.. In this study, instead, we will try to derive the usability by describing the behavior of the users; to do this we analyze their traces stored into the web server log files; these can provide an insight about the differences on the information retrievalprocess for what concerns the two kind of services that we offer (the standard web site and the mobile one); furthermore it can be considered as a quantitative method to evaluate their usability and if the last can be considered able to reduce the cost of finding and using information.

2. Materials and methods

We already made several in-depth studies on query logs related to digital libraries context, as in Scarnò (2010), or Scarnò and Sforzini (2002); in this case our data covers the date range between January and December 2012.

By following a standard analytical strategy, we first cleaned the files by retaining only the "significant" information, i.e. by considering just one of the following situations:

- the user downloaded a full text (represented by a pdf or an html file);
- the user requested a list of articles by a "search", i.e. a query in which a given term is searched in one of the fields that are associated to an article (like the title, the abstract, the author name, ISSN, author keywords or journal title);
- the user "browsed" the content of the digital library by referring to the alphabetical, category or publisher list of journals, then to the volume, the issue and, at last, to the desired full text.

Note that, according to what permitted by the web sites, operations like "bookmark", "keep me informed on new issues", and similar, generate results that belong to the browse action.

Obviously the download action is the consequence of the search or of the browsing; by following the actions made in a sequence of steps we can study the "strategy" that represents the "movements" of the users inside the digital libraries, i.e. their behaviors in the information retrieval process.

As a second step we identified the "sessions", according to Silverstein's (1999) definition: "a session is a series of queries by a single user made within a small range of time; a session is meant to capture a single user's attempt to fill a single information need."

Unfortunately specific session identifiers were not available in our files because the user can save the authentication information also for further accesses to the system. So, we used a method for session boundary detection which relies on timing characteristics, and that is based on the time between two temporally adjacent queries associated with the same user (for more details see Scarnò, 2010, cited).

After these preliminary data treatments, we verified the size of our sample and the trend of the sessions-users for the given months. Then we described some specific user'sbehaviors, like the time between two subsequent actions,

the average duration of each session and the time of the day in which the access is more frequent.

But the possibility to identify the "strategy" used on the two different web sites permitted us to build the "transition tables", i.e. frequencies tables in which the rows are related to a step and the column to the following one (where, with the term "step", it is intended one between a search, a browsing, a download).

3. Results

In the year 2012 we observed 2812 users:

- 151 (5%) accessed the digital library just by using the mobile web site;
- 1860 (66%) accessed the digital library just by using the standard web site;
- 801 (29%) accessed the digital library from both the mobile and the standard web site (we can define them as "common users").

Fig. 2and fig. 3 shows, respectively, the plot of the different accesses and of the different sessions, for each month, of the "common users" to one of the web sites (standard and mobile). These plots show a decrement in the use of the standard web site, but this result should be considered together with the increment of the accesses to the mobile web site.
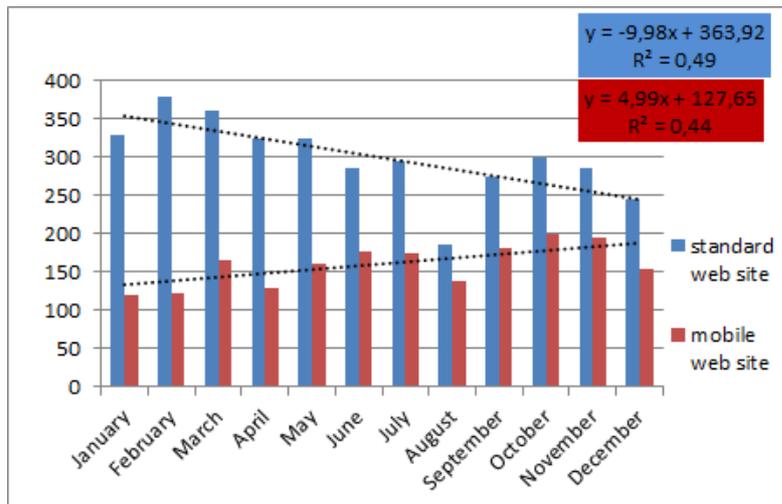

Fig. 2: plot of the different accesses, for each month, of the"common users"to one of the web sites (standard and mobile) from which it is possible to access the CINECA digital library. In the boxes there are the regression summaries

Fig. 3:plot of the different sessions, for each month, of the "common users" to one of the web sites (standard and mobile) from which it is possible to access the CINECA digital library. In the boxes there are the regression summaries
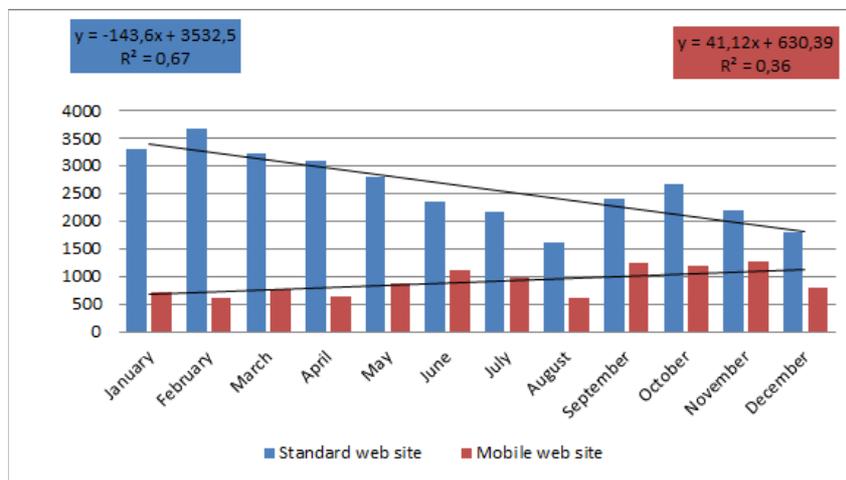
In table 1, instead, there are some summary statistics on the different behaviors that correspond to each of the two web sites; it is interesting to observe that the mobile web site, although less accessed, "captures" the users for more time; furthermore this doesn't means less actions because they do "more things" in less time.

So it seems that the mobile web site is more efficient and usable; obviously this consideration needs to be done after the analysis of what happens during a session. Table 2 and 3 contains, respectively, the transition tables between two subsequent steps made in the two different web sites.

	Accesses (mean) for each month	Duration (mean) for each session	Mean time between two subsequent steps
Standard web site	8.7	2 m. 24s.	22s.
Mobile web site	5.6	2 m. 47s	16s.

Tab. 1: summary statistics for monthly accesses and sessions according to the web site (year 2012, "common users")

Table 2 and 3 show that the most common action that the users do when entered in the digital library is a "search"; the behaviors in the two different web sites are quite similar; the main differences regard the actions related to the "search" and the download one.

In the standard web site, after a search, the results seem to be deeply analyzed, while in the mobile one the user chooses more often to make a new search.

For what concerns the download, the users of the mobile web site limit the article viewed by selecting better the results, while in the standard web site they tend to download more (maybe in order to read it after).

Standard web site				
First step	Second step			
	Principal page	Browsing	Navigation of the search results	Download of an article
Principal page	0%	36%	60%	4%
Browsing	2%	69%	4%	25%
Search	12%	12%	49%	27%
Download of an article	1%	23%	19%	57%

Tab. 2: transition table between a step and the next one, evaluated in what done by the "common users" in the year 2012 when using the standard web site to access the digital library

Mobile web site				
First step	Second step			
	Principal page	Browsing	Navigation of the search results	Download of an article
Principal page	0%	22%	78%	0%
Browsing	3%	84%	2%	11%
Search	30%	13%	37%	20%
Download of an article	4%	22%	35%	39%

Tab. 3: transition table between a step and the next one, evaluated in what done by the "common users" in the year 2012 when using the mobile web site to access the digital library

4. Conclusions

This work tries to derive the usability and the efficiency of two different ways to access a digital library. It seems that the appeal of a site studied and optimized in order to be accessed by mobile devices is growing against the one of the standard web site.

Obviously the first is less accessed, but it "captures" the users for more time; furthermore it doesn't limit to less action because the users tend to do "more things" in less time. Probably this is due to a higher simplicity of the site that can correspond to a more usability and efficiency. This consideration is, in fact, related to the observation that there are users that access the digital library only by the mobile web site (151, i.e. 5% of the registered users).

References

Borgman, C.L. (2002). Final report to the National Science Foundation. Fourth DELOS Workshop. Evaluation of Digital Libraries: Testbeds, Measurements, and Metrics. Hungarian Academy of Sciences, Computer and Automation Research Institute (MTA SZTAKI), Budapest, Hungary, 6-7 June 2002. Grant IIS-0225626

Francisco-Revilla L. et al. (2001), Managing change on the Web, *Proceedings of the first ACM/IEEE-CS Joint Conference on the Digital Libraries,* New York, ACM Press, 67-76

Jeng J. (2005) What Is Usability in the Context of the Digital Library and How Can It Be Measured, Information Technology and Librarians, June

Lesk M. (1997), Practical digital libraries: Books, bytes and bucks. *San Francisco: Morgan Kaufman*

Scarnò M. (2010), User's Behaviour inside a digital library, International Journal of Decision Support System and Technology, April-June 2010, pp. 52-59

Scarno M. and Sforzini D. (2002), La diffusione della Conoscenza via Internet, acquisizione ed elaborazione dei comportamenti degli utenti, in Data Mining, Web Mining e CRM, Franco Angeli Editore (a cura di F. Camillo e G. Tassinari)

Silverstein C., et al. (1999). Analysis of a very large Web search engine query log, *ACM SIGIR Forum*, Vol.: 33, 6--12

.