The 2nd International Conference on Integrated Information

An Evaluation Based on the Digital Library User: An Experience with Greenstone Software

Jesús Tramullas*[a], Ana-Isabel Sánchez-Casabón[a], Piedad Garrido-Picazo[b]

[a]Department of Librarianship & Information Sciences, University of Zaragoza, 50009 Zaragoza, Spain
[b]Department of Computer Science & Systems Engineering, University of Zaragoza, 44150 Teruel, Spain

Abstract

Digital libraries are a product of digital information aimed at satisfying users' information needs, through collections of documents to which value added services are provided. To ensure the correct functioning of digital libraries and the services they provide it is essential to adequately include an understanding of the needs of users and their behavior in the design and creation processes. The evaluation of digital libraries has benefited from an effort to establish theoretical frameworks. The model that is proposed is based on techniques centered on the user. The focus is on defining the structure of the analysis, establishing evaluation levels and criteria, and verifying these through an evaluation process, carried out by a group of end users. The user activity identified, and which was the subject of the study, is the process of creating, enriching and configuring a collection for a digital library. A basic evaluation instrument was designed, which combined closed and guided questionnaires with the collection and study of individual opinions of users in the test groups. The results obtained made it possible to confirm that the instrument has produced valid results. The weaknesses and problems inherent in the tool, and the tool's deficiencies which must be corrected, were identified.

Keywords: Digital libraries; evaluation; user-centered methods; Greenstone

1. Introduction

Digital libraries are a product of digital information aimed at satisfying users' information needs, through collections of documents to which value added services are provided. To ensure the correct functioning of digital libraries and the services they provide it is essential to adequately include an understanding of the needs of users and their behavior in the design and creation processes. Evaluation processes may be highly complex, given the

---

* Corresponding author. Tel.: +3576761000; fax: +3476761506.
E-mail address: tramullas@unizar.es; http://tramullas.com

numerous factors that can be involved and that must be taken into consideration [1]. This is reflected in the large number of approximations that there are, with numerous contributions to the evaluation process from different interested disciplines. From all these, it can be concluded that there is a tacit agreement that evaluation principally concerns questions that affect usability, the collection itself and the performance and features of the digital library.

Most evaluation studies pay special attention to the interface usability, and to the utility of collections and services for the users. To achieve their objectives, these studies follow user-centered methodologies [2], detailing different features of the final product. However, these analyzes do not usually take into account the processes of creation and implementation of digital collections when using a digital library creation software tool. Specifically, they pay attention neither methodological nor technological processes which are useful to collect documents, to develop metadata schemas, and to build organization and access schemes based on the content and the purpose of the collection. The functionality of both the search systems and the systems which allow accessing the information in a digital library mainly depends on the possibilities available for creating and organizing the information.

This work proposes a contextual evaluation model of the functionalities for creating digital collections currently available in existing digital library creation tools. The proposed model is based on user-centered techniques, and it focuses on defining the structure of the analysis, the establishment of standards, and the evaluation criteria. In addition, it has been tested and assessed by a group of end-users using the Greenstone digital library software creation tool [3].

2. Evaluation on Digital Libraries

Borgman [4] considered that the evolution of digital libraries has been developed from two different perspectives, corresponding to two different approaches: (i) the computer science approach, and (ii) the library science approach. At the beginning, all the efforts were concentrated on software creation tools, while later the focus shifted towards the creation of collections accounting for the end-user requirements. The literature on the subject is large, especially since 2000, and its revision can demonstrate that the evaluation process has been focused on two main areas:

E The benefits of digital libraries, from the perspective of classical algorithmic information retrieval, which is usually interested in issues regarding performance, responsiveness, reliability, accuracy, and relevance, and

E Meeting the user needs and problems related to information, focusing on the structure and organization of collections, the interfaces, and the relationships between the digital library and the goals and user mindsets.

The second area is particularly important for this work. Digital libraries, as information products designed to meet the needs of a set of users, are particularly good to be assessed by methods and user-centered techniques. In this sense, we find: (i) works that focus on the study of information content of the collection, and its relevance to the users, as well as (ii) works that focus on the analysis of the user interface, mainly centered on the analysis and evaluation of usability.

Due to the difficulty in defining models capable of evaluating digital libraries, theoretical evaluation frameworks are needed. Furh et alii [5], according to the evaluation activities developed by the DELOS network, proposed a new framework. After reviewing the currently existing approaches, mainly based on laboratory measurements such as relevance or accuracy, they claimed that new metrics were required to fit the context of digital libraries usage and the user satisfaction. Taking as a starting point the 5S DELOS digital library model, they identified an interaction model, which they coined as *Interaction Triptych*. This interaction model provides user-content relations, content-system relations, and user-system relations, which are related to usability, utility, and performance, respectively. Authors provided a set of important recommendations: (i) the need for flexible evaluation frameworks, (ii) the real users involvement, (iii) the consideration of previous experiences, (iv) the creation of a community of practice, (v) the establishment of repositories of evaluation data and format standards

for capture and processing, (vi) the evaluation of user behavior in the long term, and (vii) the attention to the specific differences in each domain of assessment. Also within the 5S theoretical model, Gonçalves, Moreira, Fox and Watson [6] analyzed the meaning of quality in digital libraries. For each concept, they proposed quality dimensions, and established numerical indicators. The measures aim to assess a final product, but neither the tools, the processes of collection, nor the information organization. Subsequently, they developed the 5SQual model [7], a quality assessment model based on the 5S model, especially designed for digital libraries. Basically, it is a tool that provides numerical values for the evaluation model indicators. Specifically, the quality dimensions assessed are: accessibility, significance, similarity, timeliness, completeness, compliance, efficiency, and confidence.

The high level of complexity that can reach the evaluation process, given the many factors that should be taken into consideration, made necessary to propose DiLEO [8], an OWL ontology able to explicitly reveal the main concepts and their relationships, as well as to serve as a support for comparative studies regarding evaluation models, and for planning future evaluations. The large number of existing approaches reflects the existence of different opinions on this field, and shows numerous contributions to the evaluation process from various disciplines. We can deduce that there is a tacit agreement that the evaluation is mainly concerned with issues affecting the usability, the collection, and the performance and benefits of digital libraries.

The 5S model is not the only approach to evaluate digital libraries. Dorward, Reinke and Recker [9] proposed a model for evaluating digital library services based on a user-centered approach, able to improve the product in successive iterations during design and development processes. To this end, they developed a survey for users, an expert review of the interface, and an advanced test prototype. Spasser [10] applied an approach based on the assessment of social realism, in order to analyze the activity theory which is applied in these processes. He established and analyzed the publication flows in the digital library studied. The evaluation focuses on reality, the context, and the elements affecting users. Xie and Wolfram [11] presented a study of end-users, based on the use of questionnaires. They improved the organization usability model, and identified three modes of interaction: influence, communication, and activity. Snead et alii [12] developed strategies for the evaluation of usability, accessibility, and functionality, by combining different methods, with an iterative user-centered study model. Khoo, in his evaluation of the NSDL [13], implemented a socio-technical theoretical framework, or "life cycle of the resource". She suggests that the evaluation can have a multi-faceted answer, and it has to address the different contexts of use of the library. To do this, the units of analysis must be identified. Creating resources and developing the collection and the metadata should be the nuclear activities to be analyzed.

In 2006, Xie [14] reviewed the existing literature on digital library evaluation. He highlighted that research efforts had moved from library construction to the satisfaction of users' needs. Therefore, to study the evaluation criteria from the viewpoint of users was necessary. Her work attempts to identify the key evaluation criteria considered by users, and the problems they face. He proposed to use the criteria established by the users themselves, which were: usability, collection quality, service quality, system efficiency, and user opinion. In a subsequent study [15], she made an analysis of the use of two digital libraries. In this analysis, users wrote down their actions and opinions in their diaries, expressing their personal evaluation criteria. Blanford et alii [16] present another advance in the user evaluation. In this work, authors focus on the user informational work perspective, putting the system in context. To do this, they create the *PRET A Rapporter* model, analyzing the purpose, resources, ethics, data collection techniques, analysis techniques, and report results. Unlike other related studies which follow classical usability evaluation models, user cognitive styles were used in the evaluation done by Frias-Martinez, Chen and Liu [17]. Recent studies have also proposed a holistic approach to set different levels of evaluation [18], or the use of participatory design techniques to create evaluation tools by using ethnographic methods [19]. In any case, it should be noted that usability concepts may significantly differ between end-users and expert evaluators, as pointed by Khoo, Kusukoni and MacDonald [20].

In the specific case of Greenstone, in contrast to the popularity and diffusion of the tool, there are few evaluation studies. Goh et alii [21] used a feature checklist to evaluate the different available free software digital

library creation tools. According to their results, they consider that Greenstone was the best tool. Blandford et alii [16] evaluated Garnet, a collection built with Greenstone and combined with a spacial hypertext system, using *PRET A Rapporter*. Their work focuses on the organizing information parameters that users manage after a process of search and retrieval.

3. Proposed Evaluation Model

The aim of this work is to establish a functionality evaluation model for the creation and implementation of digital libraries, by using a user-centered approach. Similar to the works presented in the previous section, we considered to apply end-user evaluation techniques in our proposed model. This decision was given by the objective of the analysis itself. The development of the evaluation requires the prior definition of the approach to be applied, the techniques to be used, and the development of the experiment:

1. Definition of the functionalities to be evaluated.
2. Establishment of a user activity approach.
3. Definition of the activities being performed and their control parameters.
4. Design of data collection instruments, and rating scales.
5. Execution of the evaluation process.
6. Analysis and interpretation of results.

The most appropriate evaluation approach must consider how end-users make their activities. These activities are mainly the creation process, and Greenstone's collection enrichment and configuration. One of the Greenstone's objectives is to facilitate the creation and publication of digital libraries to any user, regardless of their education profile. The activity study has been carried out in two phases:

1. Firstly, a group of users made an appraisal of the process, in response to various questions raised in a closed format.
2. Secondly, once the data from the first phase is available, a second group of users developed a second evaluation round focused on more specific issues, combining closed and open answers. This group was selected from people of the first group, attending to their greater capacities and knowledge about the object of study,

To this end, previous experiences which appear in the literature show the use of various techniques to capture data, ranging from forms to surveys, as well as focus groups, or personal diaries. In this case, we selected the use of forms that combine closed questions with open questions, and whose data can be tabulated. In addition, they can also be analyzed from an individual and more flexible perspective. Hence, a tradeoff between the adoption of closed criteria established by experts and the adoption of users' criteria is reached. The evaluation was conducted by four different groups of users, during four consecutive years, ensuring the compliance with the recommendation of evaluating in medium and long term, in order to identify the variations of user patterns that may affect the performed evaluation.

4. Applying the Evaluation Model to Greenstone

The evaluation was conducted over four consecutive academic years, i.e., during 2008-2009, 2009-2010, 2010-2011, and 2011-2012. The participants were 82 college students from degrees in both Information and Documentation, and Library and Information Science of the University of Zaragoza; 68% were women and 32% were male. All of them had received training on digital libraries and the creation of collections, as well as other related issues such as metadata, information retrieval, indexing, etc. The evaluation was presented as the final phase of a collaborative project aimed to implement a digital library using Greenstone which students implemented in two months. The Greenstone used versions were 2.83, 2.84, and 2.85. Most of the work was done using the built-in GLI (*Greenstone Librarian Interface*) Client. In a first phase, students received basic training

on the use of the tool (during two/three hours). After that, they received a document which sets: (i) the objectives, (ii) planning and project schedule, (iii) the criteria for academic evaluation, and (iv) Greenstone information support. Students had to install the tool, and then design, implement, and publish a digital library, ending with a self-assessment process which involved a closed questionnaire about the work done, and the tool used.

The questionnaire included several control questions about the development of the project, yes/no answers, and questions about valuation of supporting documentation, potential and functionality, usability and reliability, with a numerical rating scale from 1 (very low) to 5 (very high). Furthermore, the questionnaire incorporated an evaluation about the time they had spent on the project, in order to know the learning curve, and to evaluate the relationship between effort and results. The criteria and the average results obtained are shown in Table 1.

Table 1. Overall evaluation results.

Criteria	2008	2009	2010	2011
1. Use of reference documentation	3,2	3,3	3,1	2,8
2. Auto-perceived level of knowledge about Greenstone	3,4	4,0	3,6	3,2
3. Assessment of personal work	4,5	3,9	4,2	4,2
4. Assessment of support available	2,1	2,2	2,0	1,9
5. Estimation of the potential of Greenstone	4,5	4,3	3,8	3,6
6. Ease of use	2,2	2,1	1,8	1,6
7. Reliability	2,0	1,9	1,6	1,8
8. Hours devoted to the project	22,6	32,5	25,3	14,5

Users admitted they had used the reference documentation, but their evaluation was negative. After preliminary training and project implementation, they considered that they had reached an average knowledge about the tool, sufficient to satisfy the basic requirements set out in the project. In this line lies the high value they attach to their own work. It should be noted that the form also serves to self-assess the project, which may influence the estimated values by the students. The specific features of the tool, however, provided results that indicate that users considered that the tool was suffering a gradual loss of the potential to create digital libraries, and in addition its ease of use was decreasing. The low reliability that users gave the tool is important. They primarily referred to installation problems and errors not adequately explained in the creation process. This perception should be further analyzed, as Greenstone has proven to be an application that has evolved slowly but progressively, and maintains a remarkable consistency in its GLI interface, its functions, and procedures.

The second phase of the evaluation process was carried out annually on a group of six advanced users, selected according to their higher level of knowledge of the tool, which was reflected in the quality and performance of the collections that they implemented during the project. These groups completed a rating form which included sixteen questions, more specific than those included in the general form, and focused on the processes and features for creating collections, provided by the Greenstone GLI interface. The form also offered the opportunity to openly discuss each and every one of the questions raised. Table 2 shows the average numerical results obtained.

Table 2. Average results provided by the selected (second) group.

Criteria	2008	2009	2010	2011
1. Assessment of user manuals	2,0	1,4	1,2	1,0
2. Evaluation of the reference wiki	3,2	3,0	2,6	2,6
3. Evaluation of the application help	1,0	1,0	1,0	1,0
4. Ease to create the collection	5,0	5,0	5,0	5,0
5, Ease to capture documents	4	3,8	3,6	3,8
6. Ease to manage metadata schemes	1	1,8	2,0	1,6
7. Ease to assign and manage metadata	4,0	4,2	4,2	4,4
8. Ease to configure plugins	1,4	1,6	1,4	1,2
9. Ease to create indexes	3,2	3,3	3,8	3,6
10. Ease to create advanced query interfaces.	1,2	1,5	1,2	1,0
11. Ease to create simple classifiers	4,0	4,2	4,0	4,0
12. Ease to create complex classifiers	2,2	2,9	2,3	2,0
13. Ease to format the output	2,4	2,4	2,0	2,2
14. Ease to format documents	2,0	1,8	1,8	1,6
15. Understanding error messages when creating collections	1,1	1,2	1,0	1,2
16. Understanding error messages when running the tool	1.0	1,0	1,2	1,0

The indicators obtained allow to state that the reference documentation and the wiki do not conform to what users expect. The application support is unanimously considered insufficient. Some comments highlighted the lack of updated user and admin manuals, and the confusing organization of information in the wiki. These documentation problems cause a poor use of the functionalities of the tool, as noted by several users. Although the functionality and usability of the processes of collection building, and document capture and management of metadata, do not offer any problems, users find the metadata schemes management process complicated. Moreover, they consider that there is a lack of detailed information and support examples. The same deficiency was noted when addressing the question regarding the configuration of the document processing plugins that include an interface with multiple parameters and options, but with insufficient online help.

As for the 10 to 14 criteria, relating to the creation of the collection, they have a significant impact on the digital library usability. Since the elements of organization, information retrieval and access of the collection are defined by these features, the difficulty in understanding its structure and configuration bring as a result the preferential use of the simplest options, which offer lower user performance. Individual views obtained in these criteria focused on the obligation to use approximations of "trial and error" to achieve results, producing feelings of insecurity and overexertion when using the application. In addition, it was noted the lack of consistency between the plugins configuration interface, which uses different windows and detail fields (fig. 1), as well as the classifiers and formats interface, which works with large text strings with no detail (fig. 2).

Finally, the error information provided by the tool was considered insufficient. Greenstone designers opted for using a list of errors with technical information which require a certain level of knowledge of Java, XML, and

indexing for their interpretation and correction, which contrasts with the orientation toward an end-user with a medium level of knowledge about technologies, libraries, or archives, that the tool is oriented.

Fig. 1. Lack of consistency on plugins configuration

Fig. 2. Lack of detail on format configuration panel
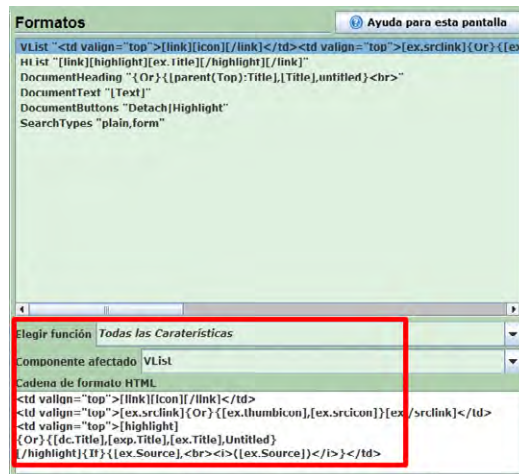
5. Conclusions

The objective of the evaluation process is to improve the product. Users have at their disposal tools for digital content management which are simplifying the process of creation and publication of digital document collections. Therefore, techniques that make possible to evaluate the technical processes of creation and implementation of digital libraries are necessary. To that end, we have designed a basic evaluation tool which combines different techniques to collect and assess the different user group views. The results obtained confirm that the instrument has provided valid results.

The evaluation was carried out focusing on the processes of creation and configuration of digital collections using the Greenstone tool. Issues related to functionality and usability have been assessed taking into account end-user problems and requirements. In addition, weaknesses and problems of the tool have been identified, as

well as the gaps that should be fixed. The problems identified during the use of the tool functionalities are caused by a low intuitive interface which presents usability problems that are exacerbated by the lack of support, and updated and properly organized reference information. The difficulty in interpreting and therefore solving the errors produced during the use of the tool should also be added. During the evaluation period, Greenstone has presented very few changes to its interface and basic features, which suggests that more noticeable changes or improvements in the interface could have changed the results. Greenstone features provided to build and publish digital libraries are sufficient and fit user expectations, so far. Usability should be significantly improved, making sure that the different elements of the collection configuration interface offer a uniform appearance and interaction. The limitations of the evaluation are given by the context in which it was made. Although user groups are future information management specialists, the environment in which the evaluation has been carried out may have introduced some bias.

References

[1] Saracevic, T. (2000). Digital library evaluation: Toward an evolution of concepts. Library Trends, 49, 350−369.

[2] Marchionini, G. (2000). Evaluation digital libraries: A longitudinal and multifaceted view. Library Trends, 49, 304−333.

[3] Witten, I.H., Bainbridge, D.,& Nichols, D.M. (20009). How to Build a Digital Library. Morgan Kaufmann.

[4] Borgman, C. (1999). What are digital libraries? Competing Visions. Information Processing & Management, 35, 227-243.

[5] Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., Klas, C.-P., Kovács, L., Landoni, M., Micsik, A., Papatheodorou, C., Peters, & Sølvberg, I.(2007).: Evaluation of digital libraries. International Journal on Digital Libraries. 8, 21–38.

[6] Gonçalves, M.A., Moreira, B.L., Fox, E.A., & Watson, L.T. (2007). «What is a good digital library?» – A quality model for digital libraries. Information Processing & Management. 43, 1416–1437.

[7] Moreira, B.L., Gonçalves, M.A., Laender, A.H.F., & Fox, E.A. (2009). Automatic evaluation of digital libraries with 5SQual. Journal of Informetrics. 3, 102–123.

[8] Tsakonas, G., & Papatheodorou, C. (2011). An ontological representation of the digital library evaluation domain. Journal of the American Society for Information Science and Technology. 62, 1577–1593.

[9] Dorward, J., Reinke, D., & Recker, M. (2002). An evaluation model for a digital library services tool. Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries. 322–323. ACM.

[10] Spasser, M.A.(2002). Realist Activity Theory for Digital Library Evaluation: Conceptual Framework and Case Study. Computer Supported Cooperative Work (CSCW). 11, 81–110.

[11] Xie, H., & Wolfram, D. (2002). State digital library usability: Contributing organizational factors. Journal of the American Society for Information Science and Technology. 53, 1085–1097.

[12] Snead, J.T., Bertot, J.C., Jaeger, P.T., & McClure, C.R. (2005). Developing multi-method, iterative, and user-centered evaluation strategies for digital libraries: Functionality, usability, and accessibility. Proceedings of the American Society for Information Science and Technology. 42.

[13] Khoo, M.: Evaluating digital libraries: A sociotechical approach. Proceedings of the American Society for Information Science and Technology. 43, 1–9.

[14] Xie, H. I. (2006). Evaluation of digital libraries: Criteria and problems from users' perspectives. Library & Information Science Research. 28, 433–452.

[15] Xie, H.I. (2008). Users' evaluation of digital libraries (DLs): Their uses, their criteria, and their assessment. Information Processing & Management. 44, 1346–1373.

[16] Blandford, A., Adams, A., Attfield, S., Buchanan, G., Gow, J., Makri, S., Rimmer, J., & Warwick, C. (2008). The PRET A Rapporter framework: Evaluating digital libraries from the perspective of information work. Information Processing & Management. 44, 4–21.

[17] Frias-Martinez, E., Chen, S.Y., & Liu, X. (2009). Evaluation of a personalized digital library based on cognitive styles: Adaptivity vs. adaptability. International Journal of Information Management. 29, 48–56.

[18] Zhang, Y. (2010). Developing a holistic model for digital library evaluation. Journal of the American Society for Information Science and Technology. 61, 88–110.

[19] Kusunoki, D.S., & Khoo, M.J. (2012). Designing digital library evaluation instruments: conceptualizing a participatory methodology. Proceedings of the 2012 iConference. pp. 632–633. ACM.

[20]. Khoo, M., Kusunoki, D., & MacDonald, C.: Finding Problems: When Digital Library Users Act as Usability Evaluators. Proceedings of the 2012 45th Hawaii International Conference on System Sciences. pp. 1615–1624. IEEE Computer Society, (2012).

[21] Goh, D.H-L., Chua, A., Khoo, D.A., Kho, E.B-H., Mak, E.B-T., Ng, M.W-M.(2006). A checklist for evaluating open source digital library software. Online Information Review, 30, 4, 360-379.