

A framework for evaluating the usefulness of digital libraries

Jasmin Hügi, René Schneider

Haute Ecole de Gestion de Genève
7, route de Drize
CH-1227 Carouge
{jasmin.hugi}{rene.schneider}@hesge.ch

Abstract

In this paper we present a methodological framework for usefulness evaluation of digital libraries and information services that has been tested successfully in two case studies before developing a corresponding tool that may be used for further investigations. The tool is based on a combination of a knowledge base with exploitable and modifiable questions and an open source tool for online-questionnaires.

1 Introduction

Recent years have seen a considerable increase in evaluating the usability of digital libraries and several methods (e.g. heuristic evaluations, user acceptance testing) have been established. Unfortunately the matter of usefulness, which has the same importance in the realm of online libraries and information services, is less investigated. It should be noted, that in the context of this paper, usefulness evaluation (of digital libraries) is seen as method, which has been developed in the context of usability evaluations and differs therefore from information quality theories and assessments.

Thus, usability and usefulness can be seen as two sides of the same coin, with a direct analogy to this metaphor, since coins can clearly be divided in an interface (usability) indicating the value that comes along with it (usefulness). The same is true for digital libraries: without a useable interface, the information behind, i.e. the digitized data is useless. On the other side, the best interface doesn't make sense if the data is not of use for the user. Since digital

libraries provide generally large-scaled masses of structured and unstructured data that generally cause high costs, the need for usefulness evaluation is evident.

Similar to usability evaluation, the usefulness of information can be evaluated before, during and after the realization of a digital library to justify the costs for new digitization or licensing or to decide to abolish existing information because users do not consider them as useful any longer.

This paper sets out to describe different methods currently in use to evaluate the usefulness of digital libraries. In the context of a research project named PECE (= Plateforme d'Evaluation pour les Centres d'Information) that started with two different field studies regarding the usefulness of digital libraries, an eleven-step approach has been developed in order to structure and facilitate future analysis of usefulness. The experiences of this study led afterwards to the creation of an electronic platform with a framework that allows an adjustable elaboration of a survey. This survey is based on questions derived from the aspects and attributes of usefulness and can be transformed to the granularity of the project under investigation. These questions will be directly transformable into a questionnaire. Persons interested in evaluating the usefulness of a digital library or its content will be able to quickly create a survey without having to consult an expert.

2 Models and methods

Interestingly, usefulness always played a minor role in human computer interaction and rarely was considered to be evaluated separately and by using specific models and methods.

2.1 Modeling

First investigations concerning a clear distinction between usability in usefulness as major factors in human computer interaction were made by Davis [89] to predict the user's intention in the so call Technology Acceptance Model based on the perceived ease of use and the perceived usefulness. Ten years later Dillon, Morris [99] developed a slightly more complex model, named P3-Framework to investigate the interdependence between power,

perception, and performance focusing on a system's utility, the user's attitude, and last but not least usability. Fuhr et al. [01] developed a first model with a specific interest for the usefulness of digital libraries with four dimensions, namely data/collection, technology, users and use and numerous measurable attributes. This concept was refined in later works, by Sandusky [02] or Saracevic and Covi [0400], who set again a focus on collecting attributes and corresponding levels that were either applied on the system or the user.

In between, Tsakonas, Kapidakis and Papatheodorou [04] and Tsakonas and Papatheodorou [06] developed a model named interaction triptych framework on which the work presented in this paper is based upon. This framework investigates the relationship between three major components: system, user, and content provided, whereas the relationship between the user and the content defines the evaluation axis of usefulness. This axis is constituted by five attributes or evaluation variables, such as a) relevance (i.e. the appropriateness of the information found corresponding to his needs); b) format (e.g. .pdf, .jpg oder .doc); c) reliability (i.e. the importance and trust given to a search result in correspondence to the information need or a given task; d) level (such as meta data, abstract, full text etc.), and e) coverage (i.e. the timeliness in terms of up-to-dateness or diachronic completeness). In a nutshell, the attributes and their equivalent variables describe the interaction between user and content as a result of the user's information literacy, his actual needs and preferences, the content at disposal and the corresponding system functionalities for presentation, transmission and storage.

As will be seen later, this model has been itself a very useful base for the investigations undergone in this project besides the fact that two more dimension have been added to cover two more aspects of a user's perception of usefulness, namely f) satisfaction, (i.e. in how far the content offered meets the user's expectations according to his own belief and desire), and, as a consequence of the latter, g) competition, (i.e. the user's competitive comparison of the system experienced after experiencing the system tested).

2.2 Methods in Use

After establishing adequate models for usefulness evaluation, the question concerning appropriate methods arises. As described earlier, the necessity for a separate evaluation of a system's usefulness or – more exactly – the

usefulness of the content provided in an information system and esp. a digital library has emerged quite lately. Hence the position or role of usefulness evaluation as opposed to usability evaluation is not really clear yet. Strictly speaking, usefulness evaluation might either be seen as a discipline *sui generis* or it can be seen as a mere offspring or sub-discipline within the overall realm of usability. The authors of this paper prefer to return to the *coin metaphor* described earlier that underlines the interdependent relationship of the two aspects and the general need to always evaluate both sides of the coin.

Along with this reflection comes the question for the specific methods that have to be used to evaluate the usefulness attributes appropriately. Following the argumentation of the preceding paragraph, purists might proclaim a methodology of its own, while others might refer to the vast number of existing and well proven methods of the usability research. Once again the coin metaphor allows a conciliatory approach, in the sense that usability testing itself might cover usefulness aspects with the help of questions covering content perception, or usefulness evaluations can be realized using existing methods while adjusting, adapting or shaping them to cover the attributes of the model described.

So far, existing studies have made use of a considerable range of methods, such as questionnaires, structured interviews, focus groups, user observation, thinking aloud tests and log file-analysis. In this context, Saracevic [04] – after having compared 80 evaluation studies – distinguishes between seven different approaches, where the orientation is either on the system, the user, the usability (as a mixture of the prior two), the ethnographic, anthropological, sociological or economic approach. Xie [06] carried out a study in which the participants could develop their own set of criteria for the evaluation of a digital library. They were then asked to evaluate a digital library of their choice by applying their criteria. In a further study, Xie [08] applied the methods of a diary, questionnaire and survey. The participants were asked to find information for six questions in two different digital libraries. They were asked to write down the whole search process. In a second part, they were asked to rate the importance of given criteria and add new ones if necessary. In a third part, the participants were instructed to apply their chosen criteria on the two digital libraries in order to evaluate them. Finally, Tsakonas and Papatheodorou [06] executed a large survey to evaluate the usability and usefulness of information services, i.e. digital libraries, electronic journals and

information portals). Besides testing different functionalities, the survey consisted of a questionnaire that set usability and usefulness into relation.

The major advantage of using a questionnaire consists in its priority given to user needs which are difficult to estimate without directly asking the user. The method also allows creating user stories or personas by either including directly open questions in the questionnaires or by abstracting them indirectly from the answers given. Both results may lead to a fact-based decision concerning the creation, deletion or adaptation of content-based services; they may also be used to define adequate user profiles for consequent usability tests.

3 Practical Implementation

As has been shown in the prior chapters, the variety of methods used for usefulness evaluations are rather low which does on the other hand not mean that every new evaluation needs its own heuristics. Besides that, there is no evidence that usefulness evaluation has to be related to some task presented to the user and thus become a different kind of usability test. To clear these circumstances, we decided to start with two separate case studies to encircle the subject of usefulness.

3.1 Case studies

The two case studies were working packages of two different, but connected research projects and commissioned by two Swiss institutions that provide information services, firstly by infoclio.ch, the Swiss portal of the historical sciences (www.infoclio.ch), and secondly by the Swiss national sound archives (www.fonoteca.ch). In both cases, all services are digital and there is no direct contact between the institution and the users.

The study concerning the Swiss historical portal infoclio.ch was built around an online survey (with incentives for participants) in order to evaluate the usefulness of the information portal. As the portal contains different databases, a blog, an event calendar etc., questions were developed specific to each component. An overall assessment done by the user was included in the survey and a few benchmark questions as well. In the end, the quantitative data collected was analyzed and presented with four personas based on this data.

The study concerning the Swiss National Sound Archives was focused around listening spots for copyrighted audio files and started with ethnographic interviews in three institutions housing a listening spot. The interviews were followed by an online survey, which contained questions focusing on the usefulness, and the usage of the listening spots and questions regarding the satisfaction of the content provided on the website of the Swiss National Sound Archives.

As stated earlier, the questionnaires for the two studies were elaborated on the basis of the attributes defined by Tsakonas and Papatheodorou [7] in their Interaction Triptych Framework (the attributes are: relevance, format, reliability, level and coverage). As it is rather rare to obtain the questions used in other usefulness studies and/or the studies contained a limited number of questions, it has been decided to develop new questions going into detail for every attribute. These questions were then discussed with the stakeholder who added new ones or deleted some of them. After discussions with the stakeholders, the model developed by Tsakonas et al. has been extended and two more attributes have been added in order to correspond to their expectations, namely satisfaction and competition.

Satisfaction was added because whenever a digital library is of use to a user and when there are no major problems with the usability, it should positively and measurably influence the user's degree of satisfaction. Therefore, the attribute of satisfaction represents an indirect way to measure the usefulness of a system. In the case of a non-satisfactory attitude of a user, it is very important to identify the source of this non-satisfaction. Is it due to usability issues and the user's inability to access a specific content although it is there? Or is it because of the non-relevance the system's content with respect to the specific needs of the user? The answers to such questions provide meaningful insight in the usefulness level of a system and indicate the components causing problems.

The attribute of competition provides knowledge about competitive services which may be consulted by a user. This attribute allows going beyond the hermetic view of a system in order to not only compare it with the users' needs and preferences, but to take its environment into account in the analysis. Any given system is always exposed to competitors, as competitive services influence the users' expectations and model their behavior.

Questions about competition allow as well comparing the proposed contents and services with similar systems. The user may then indicate preferences of each component of the system. The attribute of competition comprises as well an analysis between different components within the same system. Elements like a blog, a database or databases, a news section etc. can be compared with each other. This gives insight about the component of the system which is most useful to the users. It is then possible to adjust the invested effort to the ranking of each component.

These two new attributes cover aspects of a digital offer's usefulness which cannot be covered entirely by the other components of the Interaction Triptych Framework. Thanks to the supplementary information gained by asking questions regarding these aspects, further insights are possible. We therefore suggest to extend the existing model by these components and to take them into consideration as well when doing a usefulness evaluation.

In order to make the concept usefulness as visual and understandable as possible, a mind map was created (see: http://campus.hesge.ch/id_bilingue/doc/Usefulness_en.bmp) containing usefulness as the core concept, followed by the dimensions (or attributes), which are detailed in simple and general questions that could also show up in a survey. This concretization of parameters helped the stakeholders to understand the concept of usefulness. Besides that the mind map functioned as a starting point for further discussion and represented the central point during the whole evaluation process. As questionnaires and especially online-questionnaires are to be kept as short as possible to not demotivate the potential and actual participants, the number of questions had to be limited. The mind map allowed the stakeholders to see which aspects of usefulness could be explored and as a consequence made it easier for the stakeholders to decide what aspects are to investigate and what aspects are to ignore. In this context, it is important to know that the stakeholders were invited at several points of time to modify, add or delete questions according to the specific objectives of their evaluation. Finally, the questions were transferred into an online questionnaire, disseminated by different means of communication (e.g. mailing lists, micro-blogs, social networks, newsletters, and the stakeholder's website itself).

The evaluation process finished with the analysis of the data obtained and further explanations on them together with some recommendations for improving the service and a last meeting with the stakeholders involved in the evaluation process. For further explanations see: [Birri, Hügi, Schneider 11].

To make similar evaluations easier, the whole evaluation process, which can be interpreted as an assessment process, was subsumed in an eleven step approach as follows:

1. Brief review of the object of interest and prior usability studies regarding it
2. Analysis of the system's interface, its content and the context behind
3. Execution of ethnographic interviews to get to know the users and their needs
4. Discussion with the stakeholders about the system to investigate and about the blueprint containing the aspects and dimension under investigation (mind map)
5. Stakeholders decide which dimensions of the usefulness should be investigated with eventual additions of questions to the blueprint
6. Elaboration of the (online) questionnaire
7. Confirmation and/or modification of the questionnaire by the stakeholders
8. Survey (maybe with an incentive), including call for participation, advertisement and promotion
9. Statistical and qualitative analysis of the results
10. Compilation of the results: design of personas applying the social persona approach [Birri, Schneider 2011]
11. Presentation of the results to the stakeholders including recommendations and an impetus for their realization.

As can be seen easily, the process itself implies mutual compliance and also tries to yield compliance concerning the transfer and conversion of its result.

3.2 Building a general framework: PECE

One of the main objectives of this study was the transfer of the gained knowledge to librarians. As a consequence, the study was supposed to develop a framework which librarians can follow in order to conduct their own usefulness evaluation.

We build a model of the whole process that allowed us to extract mesoscopic questions on the base of the specific questionnaires used in each case study, which represented the most difficult undertaking. In this context, mesoscopic questions represent questions bare of every specific context like the library name or the service name. The complexity relied within the demand to keep the questions specific enough so that a) their purpose is still understandable and b) they may be adapted to every one's own context with minimum effort..

In order to make the process and the collection of mesoscopic questions available to the public, an online platform has been developed to provide a usefulness evaluation tool (usefulness.ch). This tool is based on the open source web application LimeSurvey which is dedicated to the creation of questionnaires and the conduct of online surveys. LimeSurvey is mainly based on PHP and MySQL and currently the most complete open source software concerning surveys with an active community [www.limesurvey.org].

LimeSurvey provides an enormous amount of functionalities and allows parameterizing surveys in detail. This leads inevitably to a high complexity within the interface. As many of the provided functionalities of this software were not used for the usefulness.ch platform, a new interface which works as a layer on top of the software has been created.

The platform usefulness.ch provides in addition to the online survey software a template questionnaire which contains all mesoscopic level questions. A user of this tool obtains first of all some theoretical information about usefulness evaluation and its process. After the creation of an account on the platform, the user can hence access the template questionnaire. The user can choose from this template questionnaire the questions which are to be integrated in his own questionnaire. In a further step, he may then modify the chosen questions to the context of his evaluation purposes and add his own questions. After having completed the questionnaire, it is possible to either conduct the survey directly on the usefulness.ch platform or to export the survey. The exported survey can be imported on the official LimeSurvey website (see conditions of use on their website), where the user can access the full range of functionalities if needed.

If the survey is conducted via the online platform, a link is provided which can be sent to potential participants or integrated on a website for example. The responses are collected on the platform and may be exported once the survey

has ended. The results can be downloaded in the .lsv format which is readable by any table processing program like Excel.

The usefulness.ch platform is available in German and French and its use is free of charge.

4 Conclusions

The elaboration of questionnaires, the accomplishment of related ethnographic interviews and its possible combination with other methods implies a time consuming and tedious work if done for the first time, but as soon as these methods are transferred into an equivalent framework, the same process can be done with less effort and more ease. This finding led to the creation of our knowledge base under the form of a mesoscopic collection of questions that are of interest for usefulness evaluations of digital library content. As a consequence, libraries will be able manage evaluations themselves by simply modifying the questionnaire according to their needs and accompanying them with ethnographic interviews and a log file-analysis if needed. This tool might be helpful whenever strategic decisions have to be taken, e.g. the acquisition or suppression of digital content, esp. when they imply high costs. After a usefulness evaluation, the stakeholders will have an insight in the user's needs for new services and the level of satisfaction or non-satisfaction with existing services.

Future evaluations of digital libraries will equally treat both aspects: usability and usefulness to yield a complete coverage of the user's feed-back on a digital library. However, due to the very innate relationship between the user and the content, it seems likely that this will only be possible in user centered evaluations and less with experts alone. Their task will rather consist in the elaboration of the appropriate questions.

It seems clear that after decades of putting emphasis on the evaluation of a system's usability, the exploration of its usefulness, i.e. the usefulness of a digital library's content is about to gain more and more interest. There is still much work to be done to establish this relationship and to build a solid methodological fundament. The work described in this paper shall be considered as a contribution for a further step in this process.

5 References

Birri Blezon, R., Hügi, J., Schneider, R.: "Sieht gut aus, aber was bringt es mir?": zur Evaluation der Nützlichkeit digitaler Inhalte. In: Bekavac, B., Schneider, R., Schweibenz, W.: Benutzerorientierte Bibliotheken im Web : Usability-Methoden, Umsetzung und Trends. Berlin: W. de Gruyter, 2011. S. 55-73. 2011

Birri Blezon, R., Schneider, R.: The Social Persona Approach: Using Facebook to Illustrate User Groups. In: Griesbaum, J., Mandl, T., Womser-Hacker, C. (eds.) Information und Wissen: global sozial und frei? Schriften zur Informationswissenschaft, vol. 58, pp. 327-332. VWH, Boizenburg, 2011

Buchanan, S., Salako, A.: Evaluating the usability and usefulness of a digital library. In: Library Review, vol. 58, no. 9, 638-651, 2009

Davis, F. D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3): 319-340, 1989

Dillon, A.; Morris, M.: Power, perception and performance: from usability engineering to technology acceptance with the P3 model of user response. In: Proceedings of the human factors and ergonomics society, 43rd annual meeting, Santa Monica, 1999. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.2.4868>

Fuhr, N. et al.: Digital libraries: A Generic Classification and Evaluation Scheme. In: 5th European conference, ECDL 2001, Darmstadt, Deutschland, 4.-9. September, 2001. Berlin: Springer, 2001. <http://www.sics.se/~preben/papers/ecdl-2001.pdf>.

Fuhr, N. et al.: Evaluation of digital libraries. International Journal on Digital Libraries vol. 8, 21-38, 2007

Sandusky, Robert J. 2002. Digital library attributes: framing usability research. JCDL 2002: proceedings of the workshop on usability of digital libraries at the second ACM/IEEECS Joint conference on digital libraries, Portland, Oregon, USA, 14.-18. Juli 2002: 35-38. <http://www.ucl.ac.uk/annb/docs/Sandusky35.pdf> (12.04.2011)

Saracevic, T.: Evaluation of digital libraries: An overview. Presentation at the DELOS WP7 Workshop on the Evaluation of Digital Libraries, 2004
http://comminfo.rutgers.edu/~tefko/DL_evaluation_Delos.pdf

Saracevic, Tefko; Covi, Lisa. 2000. Challenges for digital library evaluation. Knowledge Innovations: Celebrating Our Heritage, Designing Our Future. Proceedings of the 63rd Annual Meeting, November 11-16, 2000, Chicago, IL, S. 341-350.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.86.5304&rep=rep1&type=pdf> (24.06.2011)

Tsakonas, G., Kapidakis, S., Papatheodorou, C.: Evaluation of User Interaction in Digital Libraries. Proceedings of the Delos workshop on evaluation of Digital libraries, 2004

Tsakonas, G.; Papatheodorou, C.: Analysing and evaluating usefulness and usability in electronic information services. Journal of information science 32(5), S. 400-419, 2006

Tsakonas, G., Papatheodorou, C.: Exploring usefulness and usability in the evaluation of open access digital libraries. Information Processing and Management vol. 44, 1234-1250, 2008

Xie, I.H.: Evaluation of digital libraries: Criteria and problems from users' perspectives. Library and Information Science Research vol. 28, 433-452, 2006

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