Linek, Stephanie B.; Tochtermann, Klaus

Conference Paper

Sophisticated usability evaluation of digital libraries

Sophisticated Usability Evaluation of Digital Libraries

Stephanie B. Linek, Klaus Tochtermann
ZBW – Leibniz Information Centre for Economics, Kiel, Germany
s.linek@zbw.eu
K.Tochtermann@zbw.eu

Abstract: Digital libraries are an important headstone for distant education. Thereby, modern digital libraries encounter not only the classical provision of documents, but also other related services like publishing portals or lists of hot topics. Another upcoming challenge is the connection with Web 2.0. Thus, the functionalities of online libraries will change and becoming more and more complex. Accordingly, the usability evaluation of these complex functionalities has to be assured and adapted to the new challenges.
The proposed methodology of sophisticated usability evaluation follows a spiral model for prospective recommendations. The usability evaluation is done by an iterative process. This is in line with the state of the art and most modern models in usability engineering. However, in contrast to the existing models, we propose an explicit combination between systematic quantitative investigations and focused usability studies.
The proposed multi-method approach addresses different important elements of a sophisticated scientific usability evaluation. The core elements of the cyclic process are repeated usability benchmarking, focused usability studies, derivation of recommendations and decision on the planned improvements in the face of the overall strategy. The repeated benchmarking allows a quantitative measurement of advantages and drawbacks as well as improvements and impairments. The quantitative benchmarking data can also help assigning the gravidity of usability-problems or the importance of a specific innovation. Depending on the requirements of the users and the developments of the digital library, the specific usability-studies could be qualitative or quantitative or a mixture of both. The overall aim of these studies is the formulation of concrete practical recommendations. The recommendations have to be aligned with the technical possibilities and strategic decisions. After the improvements and innovations were implemented, a new benchmarking cycle can take place. The four elements can be flexible combined. This procedure ensures not only a holistic and sophisticated usability evaluation, but also the openness for new challenges and opportunities of the 21st century. The described approach will be exemplarily explained by the ZBW – Leibniz Information Centre for Economics.

Keywords/Key Phrases: Iterative Evaluation, Usability, Digital Libraries, Multi-Method Approach.

1. Introduction: Digital libraries and the importance of usability evaluation

In modern times, it's quite usual to regard to electronic literature instead of written documents. Information (including specialist literature) is more and more searched among the internet and social platforms are often used for knowledge sharing.
Even though a lot of libraries are online, many of the digital libraries appear a little bit old fashioned. In most cases, there is only a lose connection with social platforms like Facebook or Xing. Also new forms of learning and knowledge sharing (e.g., by means of serious games) are not considered. Even worse, the usability of the web-based digital libraries and its services is often problematic and usability evaluation is still an exception.
There are many slightly different definitions of usability. Besides accessibility, most of the definitions are based on four elements: Effectiveness, efficiency, usefulness (appropriateness for subjective aims of the user) and joy of use (see for example Rubin and Chisnell, 2008). The usability of Web sites in general and digital libraries in specific is of rising importance. Nowadays the end-users expect web-services, software and online platforms not only to be useful for the original purpose, but also to be self-explaining, and quick and easy to handle. In the best case, the use of software is not only helpful but also includes "joy of use".
Thus, usability evaluation is of rising importance in the modern media world. For digital libraries an adequate usability evaluation is the fundament for being capable of competing with simple search instruments (like Google) that deliver quick and easy results. While classical digital libraries offer a complete and high class literature list, many end-users prefer the "quick and dirty" search by Google, Alta Vista or similar services – just because of the better usability.
If digital libraries take their role as educational service seriously, they have to assure a good (or at least sufficient) level of usability as prerequisite to reach their target end-users.

2. Aims of usability evaluation of digital libraries

Usability evaluation should not only aim at the usability improvement of an existing online platform, but also has to account for upcoming innovations and improvements. Additionally, usability studies can also accompany the establishment of a completely new online platform or digital library (i.e., if a normal library wants to go online.)
Overall, the usability evaluation of web-based digital libraries and their services can address multiple goals:

- Identification of specific and general usability problems of existing digital libraries
- Addressing specific and isolated usability questions
- Enabling a comparison with general usability standards and comparison with similar Web sites or online-services
- Quality check of innovations and improvements: Usability validation and measurement of improvements and enhancements
- Possibility to formulate concrete recommendations for usability improvements
- Enabling and fostering the development of an online-platform of a library, that accounts for the basic requirements of usability
- Enabling and fostering the development of new online services of an existing digital library in the face of basic requirements of good usability
- Flexible methodological procedures of usability evaluations that could be applied in the case of new forms of online services, e.g., Web 2.0 applications and game-based learning

The listed goals are not unique for digital libraries. Especially, the assurance of good usability and the openness for innovations are main aims of every modern Web site and online services that have to consider new technological possibilities and the growing requirements of end-users (Dumas and Redish 1999). In this sense, the proposed methodology could also by applied for other online-services and Web sites.
It is important to note, that not all of the listed goals are of equal importance. Depending of the actual status of the Web site and online-services as well as the planned developments, different goals can be in the centre of usability evaluation. For existing Web sites, that want to incorporate innovative new applications the most important goal might be to have a flexible methodology that is apt for the usability evaluation of the new applications. In contrast, for Web sites that are still in their early development phases, a basic fundament of general usability standards could be the most important goal.
Depending of the main goal of usability evaluation one has to choose the appropriate method. For a comparison with general usability standards a quantitative standardized measurement instrument is necessary. Contrariwise, for addressing prospective usability questions of planned innovations, qualitative interview data might be more fruitful.
Nowadays there are many various methods available for usability evaluation, ranging from qualitative methods with single expert-ratings to quantitative methods with a large group of end-users. Each of these methods has its advantages and drawbacks. It has to be carefully selected which method is the best to address a specific usability problem in a specific context. Furthermore, from a practical point of view, it is also a question of resources. Even though, usability testing with end-users is often seen as the “silver bullet” for usability evaluation, sometimes it is too costly and time intensive. In literature, there is an ongoing debate which of the available methods is most efficient and effective for identifying usability problems (overview is given by Rubin and Chisnell 2008; Dumas and Reddish 1999). However, this paper is not about the comparison of different methods of usability evaluation but rather propagates a multi-method approach that is oriented on valid identification and elimination of usability problems in the face of practical possibilities.
In most cases, usability evaluation has to address more than one goal and thus, more than one single (assessment) method is necessary. Furthermore, the development of a Web site as well as the accompanying usability evaluation is a repetitive process. On the one hand, usability recommendations are needed for further developments of the Web site. On the other hand, the ongoing development of a Web site results in new usability questions.

3. Overview of the proposed multi-method approach

The proposed methodological approach accounts for the interdependence between technical developments and usability evaluation. Thereby, the need for multiple methods is explicitly addressed. Overall, usability evaluation is composed of iterative cycles with the following elements:

- Repeated benchmarking of the overall usability by means of standardized questionnaires: These data deliver quantitative indicators of improvements based on a larger panel of users
- Specific usability-studies on selected concrete usability questions based on a small sample of users: These studies work with qualitative data, but in case might also require quantitative indicators
- Recommendations based on the results of usability-studies and the quantitative benchmarking
- Improvements in the face of the usability-recommendations and strategic decisions

The iterative procedure is an important characteristic of the described approach. Usability evaluation is seen as an ongoing process that has to account for new technical possibilities as well as for changing requirements and expectations of the end users. This this is in line with the state of the art and most modern models in usability engineering (Rubin and Chisnell 2008; Dumas and Reddish 1999). However, most existing models propagate iterative testing without a specification of the measurements or how different methods could be combined in the iterative cycles.
Contrariwise, our proposed multi-method approach combines general benchmarking and specific usability studies in the face of the needed requirements and strategic decisions. In this context, it is important to note, that the whole process of usability engineering is conceptualized as an iterative cycle, including repeated benchmarking and specific usability test as well as the formulation of usability recommendations and strategic decisions.
The listed elements above could be arranged and combined in a flexible way. Thereby, quantitative and qualitative methods will be combined to enable concrete usability recommendations that can be adapted to rapid changing technology and growing user requirements.

4. Description of the single elements

4.1 Quantitative benchmarking with standardized questionnaires

Quantitative benchmarking with standardized questionnaires is necessary for a systematic ongoing usability evaluation and development of a digital library (as well as for any other long lasting professional Web service). Even though usability studies with pure qualitative data might be helpful to identify the most urgent usability problems, quantitative data with standardized measurement instruments are necessary to receive the required input for ongoing developments and a quality check of a digital library.
Quantitative usability data have the following main advantages (see also Langdridge and Hagger-johnson 2009):

- Possibility of a priority ranking of usability problems
- Quantification of improvements and enhancements (as quality check)
- Possibility of a comparison with general standards of usability
- Possibility of a comparison with competitors and alternative digital libraries (by the use of standardized existing questionnaires)

Especially the priority ranking of usability problems is of high practical relevance, because normally there are not enough resources for addressing all usability issues from the developer's side. Additionally, there might be a conflict between different aspects of usability. A classical example is the balance between simplicity of handling and the appealing design of a Web page. While qualitative data identify both problems, the quantitative measurement enables a decision for the developers: to make it either more simple (even though design will be boring) or to make the design more sophisticated (even though it will be more complicated). This illustrates also the necessity to select quantitative instruments that are apt for the specific case.
In principle, many different quantitative measurements are thinkable for usability benchmarking. Objective behavioral video data or logfile protocols as well as subjective interviews or questionnaire data can be analyzed in a quantitative way. However, the analysis of video data and logfile protocols is very resource-intensive and thus, often not feasible for a big sample or a broad variety of variables. Contrariwise, a questionnaire is timesaving and a less resource-intensive instrument. It can be easily distributed among a big sample of users; in case also online-surveys are possible.
The use of a standardized existing questionnaire allows a valid and reliable comparison with other Web sites or services and thus, enables a comparison with competitors or alternative digital libraries as well as a comparison with general standards of usability. For the assessment of general usability several standardized questionnaires are available: short ones (e.g., System Usability Scale - SUS by

Bangor, Kortum and Miller 2009) as well as long ones (e.g., ISONORM by Prümper 1999 or IsoMetrics by Gediga, Hamborg and Düntsch 1999).
However, sometimes the use of standardized questionnaires might not be sufficient since they are very general and not apt for a benchmarking of special services, e.g., the quality of the specialized content. In such cases, an additional scale has to be developed for the evaluation of specific important features, for example for the content quality of the results of a literature search.

4.2 Focused usability studies

Focused usability studies target at specific usability questions. These questions could regard to the usability of new features, to the improvement of identified usability problems, to planned innovations, to the redesign of the Web site and to many more. In this sense usability studies are aligned with the requirements of end-users, the specific questions of developers and planned innovations of the management or the general policy of the library. Accordingly, the selection of the concrete method has to be oriented on the status quo, practicability and urgency of usability issues. Thereby, many different methods are possible including quantitative as well as qualitative data. For an existing page, usability test with end users can be seen as the “silver bullet”. Contrariwise, also heuristic evaluations can be an alternative. For the establishment of a new Web site, rapid prototyping with paper-based mock-ups could be a good way (Snyder 2003). With respect to planned innovations one might prefer a focus group or semi-structured interviews. (Like mentioned above, it would go beyond the scope of this article to explain and compare the different methods.)
Each method has its advantages and drawbacks. The crucial thing is to select the most appropriate method with respect to the concrete situation, usability issue, and (last but not least) the available resources and strategic decisions. Thereby it is essential to conduct the usability study in a way that enables concrete recommendations. The “questions shapes the answers” (title of the review by Schwarz 1999). In the case of usability recommendations this implies to ask in a way that produces improvements and alternatives – and not only critique and the identification of problems.

4.3 Derived recommendations based on usability studies

Recommendations address not only the elimination of problems or the improvement of the usability in general, but also target at the accentuation of the favorite applications and the best aspects of the digital library.
Recommendations for usability improvements have to be concrete and practicable. In principle, the recommendations have to be based on the results of the focused usability studies, but also the data of the benchmarking can deliver important input (depending on the chosen questionnaires and additional questions.) Benchmarking data can be used for screening, i.e. identification of overseen or hidden usability issues. Additionally, the comparison with competitors could be helpful for finding gaps, and obstacles as well as to identify the main incentive or appeal of the own Web site.
Furthermore, also a literature research on existing usability guidelines could be a starting point, especially for the establishment of a complete new digital library or a redesign of a Web site. For more specific, concrete recommendations one has to carefully interpret the results of the own usability studies. The same is true for recommendations for planned innovations.

4.4 Making improvements: Balancing usability-recommendations and strategic decisions

In a perfect world, usability recommendations could be fully implemented and we will have happy end-users and a proud development-team. However, in reality improvements can only be made in the face of available resources and strategic decisions regarding the overall policy of the (digital) library. If resources are spare, one has to address the most urgent usability issues first. However, if there is the strategic vision to become the most innovative digital library, this might lead to the decision to address first required innovations and second the elimination of usability problems. Additionally, research-based recommendations could also contradict the policy of a digital library. For example, end-users might have the desire for a broader assortment of and easier access to the genre fantasy or horror while the management of the library wants to establish the image of an expert for sociocritical literature. It is a difficult decision, how to handle such contradictories and how to find a compromise. But think of Alice’s Adentures in Wonderland (Charles Dodgson alias Lewiss Caroll, 1865): It is sociocritical as well as fantasy literature.

5. Cyclic interplay of the four elements: adaptive iterative process

The interplay of the four elements described above can be characterized as follows:

- Multi-method approach: Combination of quantitative and qualitative methods
- Iterative cycles of usability evaluation and improvements
- Flexibility: Flexible combination of the single elements and flexibility of the single elements
- High interdependency between the elements

5.1 Multi-method approach

The combination of quantitative and qualitative methods addresses the divers demands at different stages of usability evaluation (see also Langdridge and Hagger-johnson 2009). Qualitative methods are often a timesaving and cost-effective way to identify a broad range of the existing usability problems and user requirements. However, qualitative methods do not allow a comparison with other digital libraries or search instruments. Analogously, it is not possible to quantify the improvements made in comparison with prior version of the digital library and its services. Thereby, quantitative measurements provide the necessary information. Quantitative measurements allow prioritizing usability problems as well as user requirements. Additionally, the use of quantitative standardized questionnaires enables the comparison with other software and online platforms (including other digital libraries). An important advantage of the quantitative benchmarking is the possibility to make improvements and enhancements measureable. Just the fact that a problematic feature or application was modified doesn't mean that this modification was actually an improvement. Indeed, end-users don't like changes and don't want to deviate from their ordinary handling of a Web site. Comments like "The old version was much easier" or "The old version was not perfect, but I really don't like the new one" are not an exception. Thus, changes of an existing Web site and its services have not only to be communicated, but it has also to be evaluated if the changes are actual an improvement for the end-users.
The combined interpretation of quantitative and qualitative data enables the provision of detailed concrete usability recommendations. Thereby, it is also possible to make a priority ranking between the recommendations. Based on the different methods it might be possible to make alternative suggestions how to handle a specific usability problem.

5.2 Iterative cycles of usability evaluations and improvements

The quantification of usability improvements is also a good example for the iterative character of usability evaluation and usability improvements. It is very hard to create good usability of a web-page or online service, especially for the more complex ones. Moreover, it is impossible to create an *everlasting* high usability. In line with the technological progress and the developments of the Web, e.g., the ongoing development of Web 2.0 technologies, also the requirements of the end-users will change. This in turn implies that existing Web sites and online-services have to change, too. Otherwise they will be old fashioned – or in other words they will lack usability since they are no longer efficient, effective, useful, and provide no joy of use. Thus, iterative usability measurements are necessary. On the one hand, specific usability tests have to be repeated in the face of innovations and new technical possibilities. Thereby qualitative data can be used to identify new possibilities or to identify old-fashioned applications. On the other hand, quantitative measurements and repeated benchmarking can be used as quality check if the Web site keeps up to date.

5.3 Flexibility

The iterative process is a flexible combination and arrangement of single elements. The iterative cycle can start at different points and the arrangement of the elements has to be adapted to the concrete context.
Also the single elements are flexible. Quantitative benchmarking can be made by different questionnaires and the qualitative studies can follow different approaches that should be selected with respect to the concrete practical aims and the available resources.
For an existing Web site it makes sense to start with quantitative benchmarking, to assess the status quo as a baseline and thus enable a comparison of the different versions after substantial changes have been made. Such a comparison ensures that the changes are actually an improvement for the end-users and makes improvements quantifiable.

However, for the construction of a completely new Web site one should start with small focused usability studies. Methods like rapid prototyping are advantageously to avoid wasting resources of the developer-team. As a prerequisite, the group of end-users has to be known.
In some cases also usability-recommendations could be a starting point. Until now many usability guidelines are available and could deliver fruitful input for the construction of every Web site. Even though most of the guidelines are too general to allow specific design recommendations some of them could be very helpful when open design questions come upon, e.g. how to arrange the navigation menu. A very good and valuable source of such general guidelines (based on research findings) is available at http://www.usability.gov/guidelines/guidelines_book.pdf .
Besides these rather practical, user-oriented starting points in some cases also a strategic decision could be the first step. For example, a public institution wants to invest in Web 2.0 tools or refuses the usual conventions of navigation menu to make their innovative and vanguard character visible. Such strategic decisions are a challenge for the usability researchers and require a very flexible use of usability methods.

5.4 High interdependency between elements

The elements of the approach are highly interconnected. Independent of the concrete starting point or the arrangement of elements, none of the elements can be managed isolated from the others. Quantitative benchmarking as well as focused usability studies enable concrete usability recommendations. Thereby, the measurement methods have to be carefully selected with respect to the concrete practical demands of end-users and the capacities of development. It doesn't make sense to ask the end-users for innovations that could never be implemented. Analogously, recommendations have not only to be concrete but also practicable and in line with strategic decisions.

6. Usability evaluation of the digital library ZBW – Leibniz Information Centre for Economics

In this section the described methodological approach will be explained by the concrete example of a digital library, namely the ZBW – Leibniz Information Centre for Economics (http://www.zbw.eu/index-e.html).

6.1 The digital library ZBW and its online services

The ZBW is the world's largest specialist library for economics, with locations in Kiel and Hamburg. The ZBW provides numerous services like EconBiz for literature search, EconStor as a publishing portal, EconDesk as a reference service and many more. The services are frequently used and very welcomed by the end-users. For example, for EconBiz there are about 15 000 queries per month with 20 000 to 28 000 visitors per months. Also the GooglePageRank of 7 demonstrates the high visibility and degree of popularity of this service.
Until now, the usability of the online services and the Web site itself has not been evaluated in a systematic way. Furthermore, the services will be enhanced and improved in the future. Therefore, a new internal task force for usability evaluation was established at the ZBW. The aim of this task force was not only to enhance the usability of the existing Web site and its online services, but also to regard for the future development and the planned innovations with respect to Web 2.0.

6.2 Strategic decision as starting point

The starting point for usability evaluation was a strategic decision: To become a modern digital library, that is capable to face the challenges of Web 2.0. Thereby, it was recognized, that good usability is the key to attract end-users that would be otherwise only use Google-search or similar search tools.

6.3 Benchmarking questionnaire as foundation of the usability evaluation

In a first step a benchmarking questionnaire was created, to assess the status quo of the already existing Web site and the services of the ZBW.
The benchmarking questionnaire comprises standardized existing scales, an additional scale for assessing the quality of literature search results and several prospective questions. These three main parts of the benchmarking questionnaire regard to different purposes:

Standardized questionnaires (ISONORM & SUS) serve two main aims. First, the comparability with other software products and Web sites, and second, the establishment of an internal standard, that makes improvement measureable with respect to general standards of usability.
Additional scale for assessing the quality of literature search results of the online service EconBiz tries to establish a measurement instrument for an internal quality standard of the very heart of a digital libraries online service: The literature search. The scale was constructed in close collaboration with the product manager of EconBiz and librarians. The wording and format of the scale was analogous to the SUS. Even though the validity and reliability was not proven, the scale assesses quality criteria that were judged as important cornerstones of the service EconBiz. Thereby, the scale included general criteria (e.g., completeness of the literature list) as well as specific aspects that regard to enhancements and improvements (e.g., sophisticated ranking and filter options).
The *prospective additional questions* were included to assess end-users' requirements and to receive a more holistic view. Additionally, the questions tackle also planned innovations. Besides, the end-users are asked about their typical work with the ZBW services, especially literature search with EconBiz, to receive the required information for the creation of test scenario for subsequent usability test scenarios. (More details on the benchmarking questionnaire can be found in Linek and Tochtermann 2011.)

6.4 Focused usability studies for the future development, innovations and practical needs

The usability studies of the ZBW will focus on two main areas: First, on the planned innovations based on the strategic decision (modern digital library with connection to Web 2.0 applications), and second, on the main usability issues revealed by the baseline data. These two main areas have several intersections, especially with respect to the core service of the ZBW, i.e., EconBiz, the online service for literature search. Additionally, it is planned to redesign the ZBW homepage.
The usability studies at the ZBW will mainly be done by means of videotaped usability tests. The created test scenario will be based on the typical tasks of different end-user groups. Besides usability test, the focused usability studies are designed in the face of the concrete requirements, e.g. rapid prototyping with paper-based mock-ups as information source for the redesign of the homepage and interviews with end-users as method for requirement engineering for the planned innovations.

6.5 Research-based recommendations in the light of the strategic decision

The recommendations for the improvements and further developments/innovations will be based on the benchmarking-questionnaire as well as the focused usability studies. For example, rapid prototyping and interviews with end-users will enable concrete recommendations for the structure of the sites and the labels for navigation. The quantitative benchmarking data deliver the necessary base for a priority ranking of required improvements. Additionally, the open questions of the benchmarking questionnaires provide additional information on user requirements.
After the redesign of the Web site including essential changes and innovations, the benchmarking questionnaire will be presented again to a big sample of end-users. This procedure enables a quality check, if the changes made are in fact an improvement. Thereby, the standardized usability-scales (ISONORM & SUS) as well as the scale on content quality remain constant. However, the open questions will be modified in the light of the future development and planning of the ZBW.

7. Resume and outlook

The proposed methodological approach offers a flexible way for a sophisticated usability evaluation that is on the one hand research-based and on the other hand oriented on practical needs. The combination of quantitative and qualitative data provides a more holistic view of the situation and allows the derivation of concrete usability recommendations. Additionally, quantitative benchmarking enables the comparison with older versions of the Web site as well as the comparison with general standards of usability and competitors. The flexible arrangement of elements allows a wide spread application of the methodological approach. Thereby it is important to consider the high interdependency between the elements and to adapt on the concrete situation.
The described example of the digital library ZBW demonstrates the different possibilities of the described approach. Even though the preliminary evidence (of pilot tests) is promising, it is far too early to draw conclusions yet. The described example of the ZBW is only one application case and thus, the methodology has to prove its value for other instances. Thereby, the proposed multi-method

approach and the described benchmarking questionnaire are open for adaptive variations, modifications and enrichments.

References

Bangor, A., Kortum, P. and Miller, J. (2009) "Determining what individual SUS score mean: Adding an adjective rating scale", *Journal of Usability Studies, 4(3)*, 114-123.

Dodgson, C. alias Lewis Caroll, (1865) *Alice's Adventures in Wonderland,* Macmillan: United Kingdom. Full text online available at http://www.cs.cmu.edu/~rgs/alice-table.html or at: http://ebooks.adelaide.edu.au/c/carroll/lewis/alice/

Dumas, J. S. and Redish, J. C. (1999) *A practical guide to usability testing,* Intellect Books, Oregon, USA.

Gediga, G., Hamborg, K. C. and Düntsch, I (1999) "The IsoMetrics usability inventory. An operationalisation of ISO 9241-10 supporting summative and formative evaluation of software systems", *Behaviour and Information Technology, 18*, 151-164.

Langdridge, D. and Hagger-johnson, G. (2009). *Introduction to research methods and data analysis in psychology* (2nd ed.), Prentice Hall.

Linek, S. B. and Tochtermann, K. (2011) "*Assessment of usability benchmarks: combining standardized scales with specific questions*", Proceedings of the 14th International Conference of Interactive Collaborative Learning (ICL 2011), Piešťany/Slovakia, September 2011.

Prümper, J. (1999) "Test IT: The ISONORM 9241/10", in H. J. Bullinger and J. Ziegler (Eds.), *Human-Computer Interaction – Communication, Cooperation, and Application Design* (1028-1032), Erlbaum: Mahwah, New Jersey.

Rubin, J. and Chisnell, D. (2008) *Handbook of usability testing,* Wiley Publishing Inc., Indianapolis.

Schwarz, N. (1999) "How the questions shape the answers", *American Psychologist, 54 (2)*, 93-105.

Snyder, C. (2003) *Paper prototyping: The fast and easy way to define and refine user Interfaces,* Morgan Kaufmann, San Francisco.