



THEME ARTICLE

Anthropological motivated usability evaluation

An exploration of IREON – international relations and area studies gateway

Najko Jahn

School Library and Information Science, Humboldt University of Berlin, Berlin, Germany

Abstract

Purpose – The objective of this paper is to evaluate the usability of the recently developed IREON – International Relations and Area Studies Gateway – with the aid of an anthropological motivated research design. Within such an approach, the work environment and subject experiences of the test subjects become a crucial part of the observation.

Design/methodology/approach – The objectives are achieved by contextualisation of the digital library under examination. Furthermore, previous evaluation models of digital libraries are discussed from an anthropological point of view. As a result, a multi-method approach that is context-relative and self-reflexive is applied to assess the usability of IREON.

Findings – The structural and cultural complexity of people involved in the development, operation and usage of IREON justifies a multi-method approach. Whereas information specialists and web designers tend to focus on different kind of problems, there is a high degree of common discoverability between political science students and researchers.

Research limitations/implications – Because of the contingent nature of digital library usage, evaluation methods and findings have to be always reassessed.

Practical implications – Anthropologically motivated usability evaluations are an inexpensive but efficient way to improve design activities.

Originality/value – This paper provides librarians with basic knowledge of anthropological methods to evaluate digital library services.

Keywords Digital libraries, User studies, Best practice

Paper type Research paper



1. Introduction

Librarians and information professionals always puzzle about why users do not comprehend the services they provide in a way it was intended. In the case of digital libraries, it is held that usability evaluations are best practise to solve this problem. Nevertheless, the methods available to librarians are often not sufficient because they do not take the peculiarities of digital libraries into account.

Seeing this, there remains a distinctive research problem within library and information science even if we solved all the technical problems of digital libraries. Suppose due to a usability evaluation we know that one particular digital library is effective and efficient because it takes only a short time to retrieve relevant documents. After conducting a survey we additionally know that in this case there has been a high degree of satisfaction. But do we really understand how users handle digital libraries? Do we comprehend their everyday use, their information needs and their perceptions of digital libraries? Or do these standard evaluation methods just reflect our very own conceptions and perceptions of digital library usage?

To scrutinize the impact of usability evaluations does not mean to abandon the idea of assessing the needs of digital library users, which plays an important role within the librarian profession. It is an attempt to reconsider methods in order to grasp a more profound knowledge about the pro-attitudes of someone towards a digital library. One promising strategy is to adopt methods proposed by anthropology, which has been already introduced in library and information science (Stenhouse, 1981) and applied to digital library evaluations (Crabtree *et al.*, 1997; Othman, 2004). Turning to anthropology helps librarians and information professionals to understand the services they provide better (Seadle, 2007).

Obviously, there are different ways to describe someone's attitudes towards a digital library. If we reduce it to descriptions of mere behaviour, we might fail to give a lucid explanation. But, of course, this can be done in other ways. In taking into consideration their work environment and their own subjective experiences, participants of usability evaluations become actors, whose beliefs and wants become part of the observations. We gain a thick description that Clifford Geertz (1973) posits as an epistemological maxim to act on in order to elicit meaningful data to response to certain research questions. There are indeed alternative thick descriptions possible that users might have been trying to use. In our case, the detection of the mode of dealing with a digital library as part of our professional life as well as that of users' reflects the contingent nature of every culture under examination. Additionally, librarians are always part of their own culture, which influences their interpretation of user behaviour and needs. To answer these challenges of context-relativity and self-reflexiveness, anthropologists question their empirical methods, look for alternative approaches and pay regard to subjective biases. Thus, method training is one essential part of the study of anthropology but even well-trained scholars must admit that they will not control all research methods available to them (Bernard, 1994).

Now the price might be to hefty to adopt anthropological approaches for the evaluation of digital libraries. Librarians do often have diverse backgrounds before they enter their profession. The fast technical and conceptional development of librarianship prevents many to study anthropological methods, which seem at first glance not absolutely essential but time-consuming. Yet, in this paper I will argue that methods from anthropology implicitly have always played a role in librarians' working routine. Their professional life intersects with the life of their users. To be sensitive to this fact means that the observation of library users need not to be elaborately designed any longer, since it relies on their personal experiences. Thus, such an approach is an inexpensive way to improve library services in general and the usability of digital libraries in particular.

Apart from this practical motivation, this paper examines various usability evaluation methods discussed in library and information science from an

anthropological point of view. Even though there are many techniques for collecting data about usability problems and sample characteristics, there remains a need for a strategy that maintains meaningful analysis. This is done by a multi-method approach that allows to compare findings and sample characteristics in order to achieve a holistic picture of user behaviour.

The genesis of both theoretical and practical motivation presented in this paper can be traced back to my internship at Stiftung Wissenschaft und Politik (SWP) – German Institute for International and Security Affairs[1]. SWP is an independent research institute that conducts practically oriented research to advise the German Parliament (Deutscher Bundestag) and Federal Government in foreign and security policy issues. Founded in 1962, it is now based in Berlin. With more than 130 staff members it is one of Europe's largest research institutes within its respective domain. Due to its task it seems that SWP resembles think tanks as known in the US. But the main difference is that SWP is governmentally financed and strictly party-political neutral that is reflected in its supervisory board. It includes Members of Parliament from all parties as well as Members of Ministries, academic institutes and economy.

Part of SWP is the Library and Information Service Department. It has the tasks to serve researchers of SWP, to act as information facility for Parliament and Ministries and to coordinate the German Information Network International Relations and Area Studies (FIV)[2] which is a network of the libraries and documentation departments of 13 research organisations. It aims at to maintain a national infrastructure within the field of foreign and security policy affairs.

At SWP Library and Information Service Department, I was assigned the task of developing and conducting a usability evaluation for IREON – International Relations and Area Studies Gateway[3] operated by the FIV. Apart from its own database, World Affairs Online (WAO), which includes more than 700,000 records, further databases like Public Affairs Information Service (PAIS) has been included. In order to assess which usability evaluation method to take from an anthropological point-of-view, it is a prerequisite to define the object of investigation further.

2. IREON – International Relations and Area Studies Gateway

Definitions of digital libraries serve specific purposes (Borgman, 1999). One is to distinguish digital libraries from other information services provided on the internet. According to this attempt, it is necessary to give a definite description of libraries as institutions that organize the selection, the structuring and the access to published information with regard to economical, archival and synoptical aspects; the digital library adds digital information to the classical library (Ewert and Umstätter, 1997). The other purpose is to give a conceptual model that aims not only at drawing a distinction to various information services but also revealing the diversity of digital libraries. After examining how the Digital Library Federation defines them, Shiri (2003) points out that three categories constitute the bundle of descriptions underlying digital libraries, namely people, information resources and technology. Even though digital libraries are global information systems, their use, content provision and design remain local (Duncker *et al.*, 2000), which makes the diverse nature of digital libraries evident. To illustrate this, I contextualize our digital library under investigation, IREON – International Relations and Area Studies Gateway, by means of the proposed conceptual model.

2.1 People

To differentiate at least three structural levels where people interact is a promising strategy to grasp the structural and cultural complexity of the actors involved in the development and operation of IREON. To do so, we identify decisive influence on IREON by the specialized information policy, the intra-institutional impact on research and policy consulting and the inter-institutional collaboration which broadens the information and knowledge flows among people and institutions.

In Germany scientific societies, institutions and academic subjects were served specialized information brokering, until the development of digital libraries caused policy to ask for concepts to integrate the formerly separated scientific information infrastructure in order to make it accessible to a larger community. Nevertheless, the subtly differentiated specialized information should be maintained, too. Thus, the resulting science portal *vascoda* consists of two components: separated and independent acting scientific digital libraries as well as an all-embracing portal that allows a unique access to the distributed documentary units (Krause, 2007). IREON forms part of *vascoda* portal[4].

Governmentally entrusted with the academic subject of international relations and area studies, the FIV and its members have been mainly responsible for disseminating their research institute and its researchers. But after reunification and the end of cold war, scientific consulting, for instance at SWP, broaden its scope to their respective domains. As the scientific consulting within this field is more engaged in their community, so are the connected library and information departments. Not only institute-based researchers but also scientists and students of this field have become the focus of their information services.

Governmentally financed, there still remains priority to advise the German Parliament and Federal Government. SWP Library and Information Service Department, which is in charge of FIV, keeps in close contact with Central Inquiry Point for Specialist Information of Deutscher Bundestag in order to evaluate the information needs of the Members of Parliament and to distribute them with SWP products. Central Inquiry Point for Specialist Information of Deutscher Bundestag is part of the Research Unit at German Parliament. Seven Information Research Specialists receive between 5,000 and 6,000 queries annually from Members of Parliament and internal research units. With the dependence of the information required, they either forward it to the appropriate research unit or to units of information department like the Subject and Speakers' Indexes[5] or they answer these queries themselves. For the latter purpose they refer to more than 3,000 national and international databases or to other services on the internet.

Without taking designers and content providers into consideration as proposed by Duncker *et al.* (2000), the differentiation of the levels where people interact, reveals the unique but diverse potential user community for IREON. Indeed, in an anthropologically motivated evaluation, representatives of these different kinds of potential users play a crucial role. The task cannot be only to detect potential usability problems, but also to trace them back on alleged cultural differences.

2.2 Content and technology

Alleged cultural differences are part of the information resources available and their graphical representation. For instance, the design of IREON assumes that the

researchers from the FIV institutes appreciate a prominent reference to institutional stocks. Often not accessible to external users, additional links to meta-catalogues or document delivery services are proposed to assist them. However, whether there are any conflicting design features will be only decidable after conducting usability evaluations.

Apart from people and content, technology is furthermore one aspect of the conceptional model. With regard to IREON, the integration of various databases in one digital library demands new concepts to integrate heterogeneous data. The problem of semantic heterogeneity is not only to be solved technically but also intellectual (Krause, 2007). Thus, in order to assist users not only technical but also context-related concerns have to be considered.

3. Background – towards an anthropological evaluation model

Seeing the interdependence of the aspects under examination it is easy to comprehend that usability evaluation of digital libraries is a multi-dimensional construct. The term “usability” has no clear cut definition within library and information science but has several meanings (Jeng, 2005). In her review, Jeng (ibid.) counts 14 different attributions of usability. But:

[t]he choice between the various methods is in itself a research question: the last word on evaluation methodologies has not yet been written, neither with regard to the art of methodology development nor to the craft of methodology choice for practical evaluation purposes (Fuhr *et al.*, 2007).

Therefore, this paper aims at a methodology coherent with the context of IREON that assures a self-reflexive discussion of the gathered data as well. Thus, it is helpful to adopt a first step for our conceptional framework and apply it to previous evaluation models of digital libraries before examining usability evaluation methods from an anthropological point-of-view.

3.1 Usability in the context of digital library evaluation

According to the interaction triptych model, usability relates the technical system as perceived by the graphical user interface to the pro-attitudes of the users (Tsakonas and Papatheodorou, 2006). In addition to usability, they make use of the concepts of usefulness and performance; the former concerns the relation of “user” and “content”, the latter “system” and “content”. But this seems neither to cohere with empirical findings nor with the conceptional framework proposed in chapter 2. The question to what extent the content is relevant to the user’s tasks and needs is also regarded as integral part of usability evaluations (Blandford *et al.*, 2004). Likewise Tsakonas and Papatheodorou (2006) suppose that usability and usefulness are related interdependently. To show that, they surveyed the pro-attitudes of chemists and information professionals towards usability, usefulness and performance. In result, the suggested interdependence between usability and usefulness is statistically significant.

Yet, in chapter 2, there has been a distinction between content and technology. However, in this case, content refers not only to available stocks, but considers their graphical representation in the digital library as well. Nevertheless, the contextualization of IREON shows that all three components of the conceptional

framework are related with each other. As a consequence, the usability evaluation of IREON is aware that aspects of usefulness as well as technology have to be integrated if they endanger the quality of the usability evaluations. Thus, not only the actors as proposed in chapter 2 play a crucial role in the evaluation of IREON, but information specialists responsible for the available documents and technical developer also have to be part of the evaluation in order to achieve a holistic picture of the usage of IREON.

3.2 Taxonomy of usability methods

Methods applied in usability evaluations can be subdivided into at least three different ways with regard to purpose, available user groups and empirical data-gathering methods.

3.2.1 Purpose and recourse. The distinction between formative and summative methods reflects the purpose of a usability evaluation. Formative evaluations are to inform future design activities, whereas summative evaluations provide a summary of design features (Blandford *et al.*, 2008). Many usability evaluations are formative in so far as they are motivated by the goal to achieve a holistic picture of the usage of a digital library. Such evaluations are exploratively designed to allow a wide range of possible questions and answers. Summative evaluations are the best choice, for instance, in comparison of several design features. It is also possible to combine both purposes in a study. To begin with, a formative approach elicits proposals, which can then be compared and evaluated with the aid of a summative method (Chowdhury *et al.*, 2006).

Many usability evaluations lack recourse to particular user groups. Usability inspections that involve experts in the evaluation process are a possibility to avoid this problem. Prominently introduced by Nielsen and Mack (1994), heuristic evaluation stands out with the small amount of experts who inspect a user-interface with reference to a heuristic. A task-based approach is also possible. Its benefit is that heuristic evaluation elicits many potential usability problems. However, critical voices argue that in many cases their findings do not correspond to results from user-based evaluations. Either they are too specific or they appear only once. Thus, they are often not able to give a real picture of usage (Hertzum and Jacobson, 2001). This is supported by Andre *et al.* (2003) who have found out that only 38 per cent of the detected usability problems are identical with that of the users. On the basis of a digital library evaluation, Blandford *et al.* (2004) confirm these findings. Additionally, the rate of detection differs highly between the experts involved in a heuristic evaluation, which backs Woolrych and Cockton's (2001) earlier results.

Buttenfield (1999) categorizes heuristic evaluation as informal because they can be applied in an early design process. Whereas user-based evaluations are hard to practice since the defectiveness of early designs prevents the possibility to give a holistic picture of usage, heuristic evaluation helps to focus on potential usability-problems (Hartson *et al.*, 2004). It yields a first systematic analysis, which improves the design of further evaluations applied during an evaluation process.

3.2.2 Empirical data-gathering methods. To identify a suitable usability evaluation approach depends on the decision of whether to take quantitative or qualitative research methods. If the effectiveness and the efficiency is under examination, then quantitative methods seem to be the best choice. Whereas effectiveness is best measured in terms of precision and recall, efficiency is detected by the difference between the times of request and response (Goncalves *et al.*, 2007).

Surveys and questionnaires are widely used in usability evaluation of digital libraries (Jeng, 2005). That they gather meaningful data is scrutinized by Flaxbart (2001). According to Kuruppu (2007), Flaxbart assumes a gap “between the type of questions asked in surveys, the respondents’ understanding of the questions and the investigators’ interpretation of responses”. Additionally, quantitative logfile analyses fail insofar as they miss revealing the motivation that lies behind a transaction (Nicholas *et al.*, 2006).

Seeing the limits of quantitative approaches with regard to the purpose of usability evaluations that aim wider descriptions of the mode of dealing with digital libraries, qualitative methods become increasingly relevant (Blandford *et al.*, 2008). Qualitatively gathered data provide a “rich, detailed description of activity, making use of categories that are those of the social actors themselves as they undertake their ordinary activities and make sense of the activities of others” (Crabtree *et al.*, 1997). Commonly users are asked to think aloud in order to grasp their pro-attitudes during their interaction with a digital library. Because in such a method the users verbalize their impressions, conclusions or problems, the evaluation does not only measure their behaviour but moreover describes their acting in the light of their reasons. In this way the test subjects rationalize their behaviour and make it explicit to outsiders. Thinking aloud can be provoked by tasks. Recorded comments and screen activity constitute the data basis for further interpretation.

To gain a more broader description, some refer to ethnographic methods as known within anthropology. Participating observers live within the culture they observe and take part in their test subject’s everyday life. They aim at obtaining the user’s perspective, which they make use from to improve design activities (Othman, 2004). In-depth interviews that can be open-ended, close-ended or semi-structured, are used for collecting a wealth of information (Kuruppu, 2007). Focus-groups or group interviews are appropriate if the expenses and the time are limited. However, both are open to biases because the respondents may prompt to particular answers. Because they take only a small amount of test subjects into account, qualitative approaches seem to lack representativeness (Nicholas *et al.*, 2006).

3.2.3 The anthropological perspective. It seems that both, advocates of quantitative and of qualitative approaches, were unreconcilably opposed. Indeed, even in anthropology, whose methods are mainly qualitative, there is a debate about the value of quantitative data. However, this distinction does not hold any longer if we distinguish between data gathering and data analysis. As Bernard (1996) argues, gathered data, either quantitatively or qualitatively, are not meaningful as such because they lack context-relativity. For instance, Anderson and Choudhury (2003) warn against measuring the usability of a digital library with reference to general measures. In the case of serendipity, these measures would evaluate the system as less effective and efficient. But the same holds for qualitative data gathering. In the beginning of digital library evaluation the reference to popular handbooks has been very influential because they promise usability evaluations in spite of a lack of resources.

Krug (2006) suggests only six general questions to assess the usability of a web page. Nielsen and Loranger (2006) summarized their experience in general design guidelines. Both promise the reader to deliver the best balance between revenue and expense. The presupposed universality in evaluation design and findings is applied to digital libraries fatally. It violates the diverse nature of digital libraries as shown in

chapter 2 because neither the difference to other services on the internet nor the varieties of digital libraries is taken into consideration.

Apart from context-relativity, self-reflexiveness is a further condition in anthropological research. The researcher is always bound to the own culture she or he belongs to which is reflected by his data gathering and interpretation. Contrary to Nicholas *et al.* (2006), this does not only hold for qualitative research but for quantitative research, too:

In any event, you have to talk about the text and this means you have to produce labels for themes and labels for articulation between themes. All this gets you away from the text, just as surely numerical coding does (Bernard, 1996).

A multiple choice of methods may justify or falsify findings from different evaluation phases. In the case of usability evaluation of digital libraries Craven and Booth (2006) as well as Norberg *et al.* (2005) gathered different data with the aid of different methods and compared their findings with each other.

Nevertheless, such an evaluation design seems to be too elaborate to many libraries and information departments. But there are circumstances possible under which extreme discount usability engineering produces results that respect context-relativeness and self-reflexiveness. In their approach, Marty and Twidale (2005) suggest ten minutes for the evaluator to posit context-relative questions. In a trial, they developed between one and eight tasks. Test subjects were able to answer zero to five tasks during the subsequent usability test. After that, the evaluator and the test subjects gathered feedback from an audience whether they believe that a particular event is a valid usability problem. Thus, the method provides a certain degree of self-reflexiveness, too. Even if methods are in accordance with anthropological constraints they need not necessarily violate the balance between revenue and expense. To which extent an evaluation design explains cultural phenomena decides whether the gathered data are context-relative and whether the interpretation reflects the condition of the data collection.

4. Method

In the following, the conceptional framework of digital libraries and its application to IREON as well as the taxonomy of usability evaluation methods from an anthropological point of view will be put into practice. Although the usability assessment of IREON is only one part of the evaluation of digital libraries, it nevertheless takes events into account that are beyond relationship of the system and the user. With reference to the distinction made in chapter 3, the choice of methods is made according to purpose, available user groups and empirical data-gathering methods. The chosen methods consider the anthropological requirement of context-relativity and self-reflexiveness.

4.1 Purpose and resources

The IREON usability evaluation aims at suggestions of how to improve IREON in order to be accessible to different user-groups. Therefore, it is formative insofar as it is motivated to achieve a holistic picture of usage. IREON's target user group consists of researchers of FIV institutes and their clients in Ministries and Parliament. Besides this vertical orientation the central idea of digital libraries in Germany is to enable access to

the broader scientific community including students. Thus, the other main research question is to what extent conceptions of both user groups are compatible with each other.

Particularly with regard to the general conditions of this usability evaluation, an anthropological motivated research design is appropriate. Embedded at SWP, I was able to study the everyday work routine and was able to become acquainted with the staff. Seeing this it was possible to involve researchers as well as junior research fellows in evaluation. Not only researchers and students are part of the user group, but since Information Research Specialists of Central Inquiry Point for Specialist Information of Deutscher Bundestag act as intermediary to serve the information needs of the Members of Parliament and to distribute them with SWP products, they have to be targeted, too. They are crucial actors within the context of IREON just as system developer and content provider. Due to the backing of the Head of SWP Library and Information Service Department on the one hand, and my seven month stay at SWP on the other, I was able to involve all the particular actors in the evaluation. Nevertheless, it is a prerequisite to detect appropriate methods since, as we have seen, not every usability evaluation method is suitable to every kind of usage group. Information Specialists needs may differ from the needs of researchers. As well, there might be gap between researchers and students. This is so because they work in different environments with diverse beliefs and wants towards a digital library.

4.2 Data-collection

In order to provide methods for each group of actors in the context of IREON, the proposed usability evaluation consists of several steps, namely focus-group, heuristic evaluation and user-test.

4.2.1 Focus-group. The first evaluation step elicits the information seeking behaviour of students within the field of international relations and area studies. For this purpose, seven students inform about their dealings with respective information services as well as their specific perceptions, expectations and experienced frustrations.

Information seeking behaviour is a crucial issue for a usability evaluation because it reveals demands and needs with regard to particular information services. Thus, we focused not only on electronic services but also on libraries, free search on internet and informal communication. Do students show the same or rather different strategies if seeking for information on a digital library and on the internet? Which content do they search for? Are the strategies chosen perceived as successful?

Although a presentation of IREON was not possible due to early design process the focus-group the findings were used to obtain a first impression about potential strengths and weaknesses of IREON in terms of needs for a digital library within the field international relations and area studies, content and retrieval facilities.

4.2.2 Heuristic evaluation. The second step presents a first systematic account of potential usability problems. Due to the slow development progress of IREON during that time, an expert-based method was chosen. Whereas evaluations based on, for instance, researchers and students of the field of international relations and area studies were hard to practise in an early design process, test subjects with a professional background in libraries or information departments are more appropriate. Therefore, a heuristic evaluation with five experts evaluated IREON with reference to a heuristic. Their findings were presented at a subsequent focus-group consisting of the evaluators and operators of IREON.

To reflect the different actors within the context of IREON, two Information Research Specialist from Central Inquiry Point for Specialist Information of Deutscher Bundestag participate as well as two Information Specialists from SWP who do not only supply information to researcher of SWP, but also provide content for the World Affairs Online (WAO) database which forms part of IREON database. The fifth evaluator is the webmaster of SWP homepage in order to bring technical expertise in the discussion.

4.2.3 User-test. The main component of IREON's usability assessment is the user-test for the reason that it includes the main targeted groups: researchers from FIV-institutes on the one hand and members of the broader scientific community on the other hand. The latter are represented in this study by students who aim at or who have already achieved a master's degree. A total of 12 test subjects participate, six researchers of SWP and six students. Technical developers are integrated insofar as they have been informed about the research design. They proposed further potential problems that were assessed during evaluation.

In its methodological design, user-test is task-based as well as explorative. Because tasks throw suspicion to potential usability problems as detected by the former focus-group and heuristic evaluation, they ensure comparability between the individual evaluation session with the test subjects during user-test. In conduction the interview, I beg the test subjects to think aloud. If someone deviates from the task, I do not interrupt her or him but asked for her or his reasons for his doing so. For instance, a quarter of the test subjects integrated the thesaurus in their search, although it was not part of the proposed tasks.

For every method which is task-based and explorative, there is the danger of the evaluator's biases. In order to concentrate on the interview, an open source screen-recording software records the interaction with IREON and constitute the data for subsequent analysis. This evaluation also finished with a focus-group where the findings are presented and discussed with the participants and operator.

All the methods applied, gathered rich qualitative data. To attain a more holistic picture and to avoid biases, there remains a need for a method that maintains a meaningful analysis. With reference to Craven and Booth (2006), Norberg *et al.* (2005) as well as to Marty and Twidale (2005), a multi-method approach and gathered feedback by the test subjects, the proposed evaluation design provides a certain degree of self-reflexiveness.

To achieve a even more profound understanding of sample characteristics on the one hand and usability problems on the other hand, I apply a quantitative approach for further data analysis which is also common in anthropology (Bernard, 1994). The chosen method is social network analysis (SNA), a method based on the importance of interacting units (Wasserman and Faust, 1994). For our purpose, sample characteristics and usability problems of heuristic evaluation and usability test are interrelated, in order to improves the evaluations' context-relativity and self-reflexiveness.

5. Results

The findings of usability evaluation of IREON will reveal the benefits of an anthropologically motivated multi-method approach with regard to context-relativity and self-reflexiveness. Since this paper is written for the sake of methodological

arguments, I will restrict myself to only some short remarks about detected usability problems and how they improved IREON design activities. Rather I focus on assessing the evaluation methods chosen and on sample characteristics. Furthermore, I examine to which extent the beginning contextualization of IREON and methodological discussion of usability evaluation models is justified.

5.1 Focus-group

The focus-group reveals that the participants are receptive to online information search. In their opinion, they are able to find all the relevant documents in Google, digital libraries as well as in other electronic services. They prefer direct access to documents and claim that many of the documents required are freely accessible. They are mainly interested in current literature. Even though the success of their studies confirms their search strategies, they nevertheless lack basic knowledge about offers within their domain of interest and about techniques for an effective and efficient retrieval.

These findings jeopardize two main assumptions of IREON design. On the one hand that concerns the distinction between quick and advanced search. If they are not able to refer to all the retrieval techniques available to them like indexes and thesaurus, they may miss relevant documents. In the worst case, this frustration leads to the fact that they will not rely on IREON in the future. On the other hand, due to the high affinity to freely available online contents, users of IREON may ignore the importance of bibliographic records for the scientific information retrieval.

Although the focus-group did not involve presentation of IREON, it nevertheless influences the evaluation design. It is a further purpose to evaluate to which extent the IREON homepage makes the benefits of this digital library clear to information seeking within the domain international relations and area studies. Focussing further on search and results the evaluation will reveal to which extent IREON assists users in finding their documents.

5.2 Heuristic evaluation

The heuristic evaluation shows that the homepage design is perceived as trustworthy. The evaluators focus mainly on search functionalities and results. In the former case, they criticize poorly available retrieval functionalities. It is neither possible to combine several search questions nor to make the search strategy transparent. In the case of the listed results they trouble with browsing facilities that restrict results and with the collating sequence because it uses a relevance algorithm rather than chronological order.

Apart from this, heuristic evaluation reveals many usability problems which result from IREON's early design progress. But there have been consequences for the design of IREON, too. For instance, those wording and icons that seem to be too technical or too vague are changed. But, there are also consequences for the subsequent user-test. Heuristic evaluation detects the majority of usability problems within list of results that becomes therefore a main focus of the user-test. Additionally, this evaluation proves the importance to gather feedback from the evaluator as part of a subsequent focus-group. There, further usability problems are detected and benefits and limits of this evaluation are discussed together with the participants. Seeing this, we decided to conduct a subsequent focus-group with user-test participants, too.

5.3 User-test

The user-test also confirms that the homepage design is trustworthy, so that IREON is at first glance a relevant service to find specialized information within the domain of international relations and area studies. Seven out of 12 test-subjects start to refer to advanced search that is a rather interesting fact because only students and researchers participated during the user-test. Nevertheless, they have not been able to make sense of all the possibilities IREON has to offer for a successful advanced search. Even though the wording “thesaurus” were changed to “keywords” after heuristic evaluation, its functionality remains unclear to all the test subjects. Thus, it confirms the focus-group’s finding that many of the participants who are not information professionals lack knowledge of retrieval techniques. Browsing facilities to restrict results remain a worry since only a small amount of test subjects perceived this possibility. Those who utilize them regarded them as very useful.

The user-test also confirms the high affinity to directly accessible documents. But if direct access is not possible, the test subjects resort to additional hints how to get their document, which, in turn, invalidates worries after the focus-group. It reassert heuristic evaluation insofar as icons are still misleading. A further result is that the test subjects prefer the collating sequence to be in chronological order.

IREON assumes that researchers from FIV institutes appreciate a prominent reference to institutional stocks. Often not accessible to external users, additional links to meta-catalogues or document delivery services are proposed to assist them. Most test subjects understand this distinction. The application of SNA will reveal that even though test subjects were divided according to their target user group, the sample students on the one hand and FIV-based researchers on the other hand reveal a common discoverability.

6. Discussion

During evaluation of IREON, findings from different evaluation phases are compared with each other in terms of quantified sample characteristics and usability problem distribution.

The heuristic evaluation justifies to include both information professionals of SWP and Deutscher Bundestag because as we see in the left of Figure 1, the network is divided into three components. This can be explained with reference to the participants. The component on the upper left represent the usability problems observed by the Information Research Specialist of Central Inquiry Point for Specialist Information of Deutscher Bundestag, the component on the lower right by the Information Specialist of SWP and that on lower left by the web-designer. Therefore, the needs and perceptions of IREON differ between Deutscher Bundestag’s and SWP’s information professionals, which can be explained by their distinctive professional demands. Additionally, it reveals that the web-designer focuses more on general usability problems than on design features which are essential for specialized information services. This is so because the web-designer criticized only two out of 26 usability-problems in list of result.

Contrary to that, usability-problem distribution of user-test reveals a more homogeneous picture which shows the right network in Figure 1. Usability-problems interrelated via researchers and students form one component. Since information professionals and web designers tend to focus on different kinds of problems, the

LHT
26,4

618

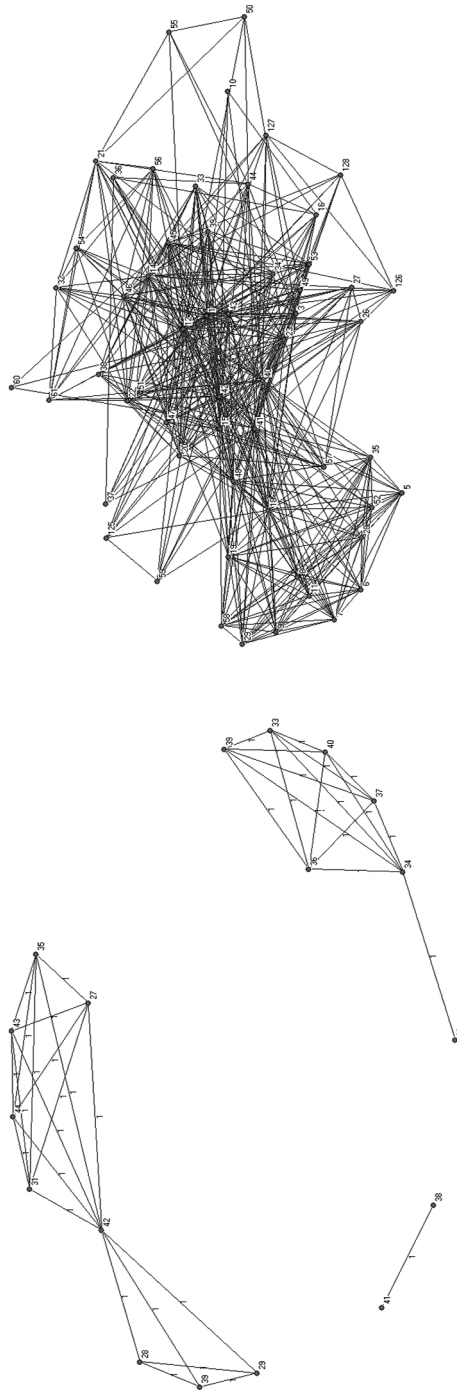


Figure 1.
Distribution of usability
problems within the scope
“list of results”

quantification reveals the high degree of common problem discoverability between students and researchers. Thus, the application of a quantitative method proves in that case the distinction between researchers and students wrong. Both are part of the very same culture with regard to the usage of IREON.

7. Conclusion

An anthropologically motivated usability evaluation does not only rely on contextualization of the digital library under examination, but assesses the condition underlying the evaluation as well. In doing this, it does not look up for an elaborate and expensive evaluation design as proposed by professional usability laboratories. Rather it makes uses of resources already available that are manifested in the intersection with the life of users. In involving different kinds of actors in usability evaluation, asking them about their beliefs and wants and observing their working routine, anthropology helps librarians to understand the services they provide in a inexpensive manner.

Of course, as we have seen, as a matter of fact such an approach confronts us with a vast amount of data that do often not cohere. A multi-method approach allows for comparison insofar as it either justifies or falsifies earlier findings. In addition, quantitative methods like SNA do not deny the essential role of qualitative methods of data collection on usability evaluations. But with its data, it provides new insights into the usability evaluation. It is a further step to gain a holistic picture of the usage of a digital library as well as to reconsider the applied evaluation design.

Indeed, if we accept the contingent nature of digital libraries, puzzling about users perceptions of library services will never end. But if librarians and information professionals continuously reassess and improve evaluation methods and design activities, they shall diminish this problem in future.

Notes

1. www.swp-berlin.org/
2. www.fiv-iblk.de/eindex.htm
3. www.ireon-portal.eu/
4. www.vascoda.de/
5. www.bundestag.de/htdocs_e/documents/subject.html

References

- Anderson, T. and Choudhury, S. (2003), "A usability research agenda for digital libraries", available at: http://dkc.jhu.edu/usability/_1.html (accessed 25 March 2008).
- Andre, T.S., Hartson, R.H. and Williges, R.C. (2003), "Determining the effectiveness of the usability problem inspector: a theory-based model and tool for finding usability problems", *Human Factors*, Vol. 45 No. 3, pp. 455-82.
- Bernard, R.H. (1994), "Methods belong to all of us", in Borofsky, R. (Ed.), *Assessing Cultural Anthropology*, McGraw-Hill College, New York, NY, pp. 168-77.
- Bernard, R.H. (1996), "Qualitative data, quantitative analysis", *The Cultural Anthropology Methods Journal*, Vol. 8 No. 1, pp. 9-11.

- Blandford, A., Adams, A., Attfield, S., Buchanan, G., Gow, J. and Makri, S. (2008), "The PRET A Reporter framework: evaluating digital libraries from the perspective of information work", *Information Processing and Management*, Vol. 44 No. 1, pp. 4-21.
- Blandford, A., Keith, S., Connell, I. and Edwards, H. (2004), "Analytical usability evaluation for digital libraries: a case study", *IEEE Computer Society & Association for Computing Machinery*, Association for Computing Machinery, New York, NY, pp. 27-36, Proceedings of the 4th ACM/IEEE Joint Conference on Digital Libraries.
- Borgman, C.L. (1999), "What are digital libraries? Competing visions", *Information Processing and Management*, Vol. 35, pp. 227-43.
- Buttenfield, B. (1999), "Usability evaluation of digital libraries", *Science and Technology Libraries*, Vol. 17 Nos 3/4, pp. 39-59.
- Chowdhury, S., Landoni, M. and Gibb, F. (2006), "Usability and impact of digital libraries: a review", *Online Information Review*, Vol. 30 No. 6, pp. 656-80.
- Crabtree, A., Twidale, M.B., O'Brien, J. and Nichols, D.M. (1997), "Talking in the library: implications for the design of digital libraries", *DL '97: Proceedings of the Second ACM International Conference on Digital Libraries*, ACM, New York, NY, pp. 221-8.
- Craven, J. and Booth, H. (2006), "Putting awareness into practice: practical steps for conducting usability tests", *Library Review*, Vol. 55 No. 3, pp. 179-94.
- Duncker, E., Leng Theng, L. and Mohd-Nasir, N. (2000), "Cultural usability in digital libraries", *Bulletin of The American Society for Information Science*, Vol. 26 No. 4, available at: www.asis.org/Bulletin/May-00/duncker_et_al.html (accessed 25 March 2008).
- Ewert, G. and Umstätter, W. (1997), *Lehrbuch der Bibliotheksverwaltung*, Hiersemann, Stuttgart.
- Flaxbart, D. (2001), "Conversations with chemists information-seeking behavior of chemistry faculty in the electronic age", *Science and Technology Libraries*, Vol. 21 Nos 3/4, pp. 5-26.
- Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P. and Kapidakis, S. (2007), "Evaluation of digital libraries", *International Journal on Digital Libraries*, Vol. 8 No. 1, pp. 21-38.
- Geertz, C. (1973), "Thick description: toward an interpretive theory of culture", *The Interpretation of Cultures*, Basic Books, Inc., New York, NY, pp. 3-30.
- Goncalves, M.A., Moreira, B.L., Fox, E.A. and Watson, L.T. (2007), "What is a good digital library? A quality model for digital libraries", *Information Processing and Management*, Vol. 43 No. 3, pp. 1416-37.
- Hartson, R.H., Shivakumar, P. and Perez-Quinones, M.A. (2004), "Usability inspection of digital libraries: a case study", *International Journal on Digital Libraries*, Vol. 4 No. 2, pp. 108-23.
- Hertzum, M. and Jacobson, N.E. (2001), "The evaluator effect: a chilling fact about usability evaluation methods", *International Journal of Human-Computer Interaction*, Vol. 13 No. 4, pp. 421-43.
- Jeng, J. (2005), "Usability assessment of academic digital libraries: effectiveness, efficiency, satisfaction, and learnability", *Libri*, Vol. 55, pp. 96-121.
- Krause, J. (2007), "The concepts of semantic heterogeneity and ontology of the semantic web as a background of the German science portals vascoda and sowiport", in Prasad, A.R.D. and Madalli, D.P. (Eds), *International Conference on Semantic Web & Digital Libraries (ICSD 2007)*, pp. 13-24.
- Krug, S. (2006), *Don't Make Me Think: A Common Sense Approach to Web Usability*, 2nd Ed., New Riders Press, Indianapolis, IN.
- Kuruppu, P.U. (2007), "Evaluation of reference services: a review", *The Journal of Academic Librarianship*, Vol. 33 No. 3, pp. 368-81.

-
- Marty, P.F. and Twidale, M.B. (2005), "Extreme discount usability engineering", GSLIS UIUC Technical Report, available at: www.isrl.uiuc.edu/~twidale/pubs/ExtremeDiscUETechReport.pdf (accessed 25 March 2008).
- Nicholas, D., Huntington, P., Jamali, H.R. and Watkinson, A. (2006), "The information seeking behaviour of the users of digital scholarly journals", *Information Processing and Management*, Vol. 42 No. 5, pp. 1345-65.
- Nielsen, J. and Loranger, H. (2006), *Prioritizing Web Usability*, New Riders, Berkeley, CA.
- Nielsen, J. and Mack, R.L. (1994), *Usability Inspection Methods*, John Wiley & Sons Inc., New York, NY.
- Norberg, L.R., Vassiliadis, K., Ferguson, J. and Smith, N. (2005), "Sustainable design for multiple audiences: the usability study and iterative redesign of the documenting the American South digital library", *OCLC Systems & Services*, Vol. 21 No. 4, pp. 285-99.
- Othman, R. (2004), "An applied ethnographic method for evaluating retrieval features", *The Electronic Library*, Vol. 22 No. 5, pp. 425-32.
- Seadle, M. (2007), "Rezension zu: Petra Hauke; Konrad Umlauf. Ed. (2006) Vom Wandel der Wissensorganisation im Informationszeitalter", *LIBREAS – Library Ideas*, Vol. 3 Nos 1/2, available at: www.ib.hu-berlin.de/~libreas/libreas_neu/ausgabe8/014sea.htm (accessed 25 March 2008).
- Shiri, A.A. (2003), "Digital library research: current developments and trends", *Library Review*, Vol. 52 No. 5, pp. 198-202.
- Stenhouse, L. (1981), "Using case study in library research", *Social Science Information Studies*, Vol. 1, pp. 221-30.
- Tsakonas, G. and Papatheodorou, C. (2006), "Analysing and evaluating usefulness and usability in electronic information services", *Journal of Information Science*, Vol. 32 No. 5, pp. 400-19.
- Wasserman, S. and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, MA.
- Woolrych, A. and Cockton, G. (2001), "Why and when five test users aren't enough", in Blandford, A. and Vanderdonckt, J. (Eds), *People and Computers – Interaction without Frontiers*, Joint Proceedings of HCI 2001 and IHM 2001, Vol. 15, Springer, London, pp. 105-8.

Corresponding author

Najko Jahn can be contacted at: najko@gmx.de