



Cover Page for Take-Home Assignments/Synopses For Assignments at the Faculty of Humanities

Personal information		Department
Name		<input type="radio"/> ENGEROM
Student email (e.g. abc123@alumni.ku.dk)		<input type="radio"/> IKK
Phone number		<input type="radio"/> INSS
		<input type="radio"/> IVA
		<input type="radio"/> MEF
		<input type="radio"/> SAXO
		<input type="radio"/> ToRS

Study information	
Study programme	
Level	<input type="radio"/> BA <input type="radio"/> BA elective <input type="radio"/> KA <input type="radio"/> MA elective <input type="radio"/> Open University

Information about the exam			
Title of course			
Title of exam			
Subject-element code			
Curriculum			
Examination period			
Assignment title			
Standard pages	No. of characters	May the assignment be incl. in the assignment library?	<input type="radio"/> Yes <input type="radio"/> No
Examiner			

Information in case of group examinations			
The author of each section in the paper is clearly stated (tick off)		<input type="radio"/> Yes	<input type="radio"/> No
Co-writer 1	Name		KU username
Co-writer 2	Name		KU username
Co-writer 3	Name		KU username
Co-writer 4	Name		KU username

Declaration of Academic Honesty

I hereby confirm that I have completed this assignment on my own. In case of group examination, is the author of each section in the paper clearly stated in accordance with rules on group examinations. All quotations in the text have been marked as such and the assignment or fundamental parts of it have not been presented before in other contexts of assessment. The maximum number of standard pages has not been exceeded. *Handing in the assignment electronically by logging in at Absalon and uploading the material will replace a handwritten signature.*

You can find instructions on how to merge two PDFs on <http://absalon.hum.ku.dk/english/students/onlinesubmission/>

Abstract

Purpose – In order to connect the various international research hubs on 'physical learning spaces', a large-scale research database has been developed, using a Zotero group. Hitherto, its interface and collection index has never been examined for usability. This pilot study attempts to discover what retrieval strategy combinations users apply in the Zotero web interface, and how satisfied they are with the usability and the retrieval outcomes. The results shall not just generate ideas for the improvement of the studied database, but also provide inspiration for similar Zotero projects.

Design/methodology/approach – This pilot study is designed as a qualitative field study. A sample of the project's actual target group was contacted around Copenhagen, Denmark. During a home- or office-visit, a natural search task was defined and executed by the participant on a laptop provided by the instructor. Using TechSmith's Morae usability software, screen, webcam, and voice data was recorded and analyzed; after the recording, a usability survey was filled out.

Findings – Despite only two samples, the participants use and judge the three search methods of Zotero differently. Most participants favor the free text search method (1), although the retrieval results are unsatisfactory. In a large-scale, multi-language collection, like the assessed database, browsing in hierarchical categories (2), or faceting results using a tag cloud (3) may be more effective and efficient, but only a minority of participants understands and applies these methods. Furthermore, it appears that the interface lacks intuitive navigation, especially for the non-scientific community. Novice Zotero users not familiar with the concepts of bibliographic databases may fail to differentiate between the Zotero website (the service provider) and the Zotero group (the database, the actual subject of the study).

Originality/value – This is the first published usability study of a large-scale Zotero group. It introduces usability issues, regarding search functions and web interface. Besides drawing inspiration from a similar Zotero bibliography, which uses RSS feeds and API interfaces, a few practical ways to enhance user search experience are suggested. The pilot study concludes with suggestions for further research, designed for more reliable participant scales.

Figures and tables

Figure 1: The screen setup after the Morae recording is started.....12

Table of Contents

Abstract	1
Figures and tables.....	2
Table of Contents.....	3
Introduction	4
Literature review	5
A similar large-scale Zotero group project	5
Methodology	7
Methodology development and discussion	7
The final experimental setup.....	8
Results	9
Participant 1 (P1).....	9
Participant 2 (P2).....	9
Participant 3 and 4 (P3 and P4)	10
Data analysis	11
Discussion	13
Differentiating between the Zotero website and the 'Zotero group'	13
Assessing the three search methods for user satisfaction	13
Conclusion	14
Perspectives for further research.....	15
References	17
Appendix.....	19
Morae manager error message for recording of P2.....	19
Morae post-recording survey	20
Statement of informed consent	23

Introduction

Since spring 2013, the author is building and curating a large-scale bibliographic research database on physical learning space development ("Learning Spaces - Lernorte," 2014). The project is funded and supervised by Dr. Prof. Richard Stang from Media University Stuttgart. It intends to map and connect the various national research hubs, and thereby provides guidance and best practice in the changing field of learning space – especially in Germany, among libraries, adult education centers, and schools. The literature collection consolidates all relevant publications and projects in Germany and most of the English-speaking world since 2006. Bibliographic management software Zotero was chosen as the most cost effective solution for managing a collection, while having it publicly accessible on the web as a "Zotero group". Hitherto, the database contains 1470 documents and will grow steadily. An official release is scheduled for Summer 2014.

Throughout the curation, none of the actual target group has ever had a look at the collection and its structure; never did the author get any feedback on how sensible the presentation and the applied indexing strategies were. In order to see if the current structure can fulfill users' special information needs, this paper examines the outcomes of a *pilot usability field study*, using a *sample of participants of the project's target group*, and *personal, natural search tasks*. As a methodic framework, TechSmith's *Morae* software is used, in order to capture a maximum amount of usability data, while allowing participants to carry out the tasks in their natural work environment.

While the interface of the Zotero web service is static, there are three search strategies that a user can apply to retrieve bibliographic data:

- 1) *Key word search* in the search bar on the top right¹
- 2) Browsing in the *classificatory folder structure* on the left hand side
- 3) Searching in the *tag cloud* below the classification structure²

Using natural search tasks, the study endeavors to discover *what retrieval strategy combinations users apply in the Zotero web interface, and how satisfied are they with the usability and the retrieval outcomes*. Through satisfaction measure and qualitative feedback, the user friendliness and usability issues are uncovered.

¹ There are two modes of key word search: either searching in *all indexes*, or in the "*Title, Creator, Year*" indexes.

² The tags are mirrored the classification structure, which means that *every class has a synonymous tag*. The tags form a controlled vocabulary, which can be combined in Boolean logic. Also, every specific institution dealt with in a document is assigned a tag.

Literature review

Scientifically, Zotero is dealt with only marginally. There is a large amount of instruction and bibliographic-management-software comparison websites and literature, however, rarely with critical or evaluative foci on the interface. Lauterbach et al. (2009) contribute the only usability study on Zotero; its scope varies from this paper's research design, since they assess the *standalone software*. Hence, no triangulation advantage can be taken from their findings, since this study addresses the web interface and only specific, natural search tasks. On the contrary, their applied methodology is inherently similar to the one initially planned and applied by the author in this study.

It is assumed that the web interface of Zotero groups has not yet been assessed in a usability study – or at least, no results have been published. Despite this knowledge gap, there exist a number of large-scale collaborative bibliographic databases that use Zotero to display their content (for example "Ancient World Open Bibliographies," 2013; Hiltmann, 2014). Hence, the author approached the responsables behind some of these groups, in order to probe, if they have evaluated the usability of their collections and the interface.

A similar large-scale Zotero group project

In 2011, Franziska Heimbürger initiated a collaborative bibliography for the *International First World War Studies Society* (IFWWSS) ("First World War Studies Bibliography," 2014). The database covers nearly 4000 items and 46 group members, potentially contributing to the bibliography (while Heimbürger and four others are mentioned as the major contributors in adding new references ("Collaborative Bibliography," 2014)); the Zotero group is 'open', so that anyone can join and add on to it. Heimbürger, in direct correspondence (2014), but also in various talks project talks (2013a, 2013b, 2013c), stated that the classification structure used is a slightly modified version of Horne's taxonomy in his *Companion to World War I* (2010) – only "a couple of sub-categories" were added. It remains flat, with only two levels of depth. Tags are only used to facet document language and minor attributes. From the RSS functions that Zotero supports, the collection is spread (and marketed) through Twitter and a Facebook profile³.

The classification (and this is the innovative aspect of this project) is displayed on ISFWWS's website ("Collaborative Bibliography," 2014), using the Wordpress plugin "Zotpress"; this means, there are two portals and interfaces to access the collection. On the society's website, every classification descriptor is displayed as a tile; selecting it loads the bibliographic data of the related entries in plain text. The society still recommends the Zotero web interface to "search across the entire content and filter columns by author or title" and eventually suggests to join the Zotero group, synchronize the database with the standalone software of Zotero, in order for researchers to process the data with the functions they are 'used to' ("Collaborative Bibliography," 2014).

While the above measures resulted in a success of the collection, in terms of growth⁴, there has been no feedback mechanisms or usability testing before or after implementation (Heimbürger, 2014)⁵. It appears that only qualitative remarks by colleagues were used as indicators. In total, this is the same situation of this study's subject ("Learning Spaces - Lernorte," 2014), although Heimbürger's bibliography is of a larger scale, concerning any dimension, especially technology implementation. Consequently, also for collaborative bibliographies like this one,

³ „I see a fair amount of interaction through these means: people adding @fwwbib to their favorites, retweeting them, sometimes mentioning a colleague who would be interested, and also likes and comments on Facebook" (Heimbürger, 2014)

⁴ In the first nine months, the collection grew 62% - especially during the summer period (Heimbürger, 2013a, 2013c); as of today, it has more than doubled.

⁵ "Otherwise feedback is more sporadic - things like colleagues I meet at conferences mentioning how useful they and their students find the bibliography as a resource. There is currently no way of getting analytics feedback on usage of the web view of the Zotero bibliography, so I don't have figures on that." (Heimbürger, 2014)

concrete usability results would make a significant contribution to evaluate collection structures, interface design, and eventually, the cost-efficient allocation of project resources.

Methodology

Inspired by Heimburger's work, and in context of the research question, this study's alternative hypothesis (based on Lazar, 2010, p. 23ff.) is:

H₁: *There is a difference in user satisfaction between the three search methods available in the Zotero web interface.*

H₂: *The satisfaction with the used search methods in the Zotero web interface is depending on the index structure of the information objects⁶.*

As a pilot study, this hypothesis is assessed through qualitative, descriptive data (Lazar, 2010, p. 21). The major *independent variables* are:

- The Zotero web interface (including the three search methods)
- The information objects of the database and their index structure

These elements are measured through the following *dependent variables*, based on (Jordan, 1998, p. 18f.):

- The participants' *satisfaction* with their retrieval results and the interface design
 - o Their perceived *effectiveness* of the retrieval and the search methods available
 - o Their perceived *efficiency* of the available search methods

In order for these variables to remain reliable and valid, the following variables have to be *controlled*:

- The participants need to have an *interests in physical learning spaces*, and in scope of the project's target group
 - o A valid natural search task focuses on an apparent information need of the participant, or at least an interest in relation to their work or research field.
- The test has to be carried out *in the participants' natural work/research environment*, in order to qualify as an 'ecologically valid' (Jordan, 1998, p. 63) field study.
- The *briefing and the search task definition* has to be carried out *in the same manner and without revealing information about the database platform, the structure, or the collection*. Otherwise, the test would not resemble the intended field scenario (a participant visiting the database for the first time).
- The hardware and software configuration has to be the same at all tests.

Methodology development and discussion

TechSmith's *Morae* is a software that records screen captures, microphone and webcam input, designed to track users' navigation behavior for usability purposes. There are no constraints of having actual hardware placed on one's head, or being restricted to a laboratory space. A software-based method adds flexibility to the research design, which enables the study to be carried out in other places than the laboratory, reducing the amount of hidden variables of the experiment (while adding others (Jordan, 1998, p. 63f.)).

Originally (and ideally), the author intended to use *Morae*, *installed on the participants' regular workspace computers*, in order to minimize distracting variables that would alter the state of their natural work environment. Unconfined to physical space, a multitude of international researchers and practitioners would qualify for the study, yielding a large and broad sample of the project's target group – the population. However, it turned out that this is an unrealistic study design, because of the long preparation and installation time for the participant. A

⁶ This hypothesis is not just relevant, but necessary to impose, because the way an information item is structured in Zotero determines the search methods available to the user, in the end. (2) and (3) are a direct result of indexing an information object in a Zotero "collection" or assigning a tag to them.

preliminary dry run of the installation and recorder utility was accompanied by manifold smaller distractors and major errors, which required the adjustment of system settings. In short: the participants would have been asked to install a software on their private computers, taking into account the risk of an unsuccessful installation, which might sabotage or prevent the recording of data. In turn, the participants' satisfaction would very likely be affected by these complications, as well. This scenario would be a *fatal error* (Jordan, 1998, p. 21).

The best compromise, combining natural field study benefits and the unobtrusiveness of Morae was to have the recorder software installed on a laptop (the author's personal one), brought to the participants' natural work environment, for them to run the test on the device. The participants are no longer required to prepare a test environment themselves – it simply comes to them. While this 'on-site' research design is praised by TechSmith for its cost-efficiency and mobility (TechSmith Corporation, n.d.), the design introduces a number of conflicting elements: Alien hardware and software is introduced into the participants' work environment. Every participant is used to different input devices, browsers, screen sizes, interface languages, and software tools, like citation management software (for example, P1 has to resort to pen and paper at 10:31), yield distractors that may influence the naturalness of the test environment.

Elevating this study to a possible 'home-visit' raises ethical considerations towards the participants (Courage & Baxter, 2005, p. 95ff.). Entering their (possibly private) workspace and recording parts of their information seeking behavior calls for a recording and use permission, as well as anonymity (using participant IDs, e.g. P1, P2, etc.), the right to withdraw, and the general right to be informed about the processes and intentions of the study. Hence, a "statement of informed consent" (after Courage & Baxter, 2005, p. 98f.) has to be signed by the participant before the test commences (see appendix).

The final experimental setup

In the Morae recorder, default values and options are adjusted to prevent obtrusiveness when running in the background of the laptop, and the post-recording survey is prepared. To find participants, the database itself can be consulted to find Danish authors, which are afterwards contacted with the incentive of contributing to a major international database in their field. Upon success, a date and a location for the field study is agreed upon. On-site, after a short briefing, the natural search task is found by a quasi-reference-interview of participants' knowledge and practice background and by probing for current challenges that they or the field face⁷. Including debriefing and deconstruction, a home-visit should not take longer than an hour to an hour and a half (Lazar, 2010, p. 52).

At the meeting, the laptop and the recorder software is set up at the usual workspace of the participant. Last clarifications and practicalities are settled, the recorder and the task is started and the research conductor leaves the room. The participant is told that the data will only be used anonymously, and will eventually be analyzed and deleted after the evaluation of the entire assignment. They may break off the study at any point, and are not restricted to use other web tools and search strategies, if they feel dissatisfied with the database.

Upon ending the recording, a 'post-recording survey' will pop up from Morae. The survey aims to probe for the usability measures of the retrieval *efficiency*, its *effectiveness* compared to the participants' usual research methods, and foremost, their *satisfaction* (Jordan, 1998, pp. 18ff., 65ff.). As it is difficult – even for the author and curator of the database – "to anticipate what the usability issues will be" (Jordan, 1998, p. 63), open-ended

⁷ It makes sense to call this pre-study activity an 'unstructured interview' (as in Lazar, 2010, p. 189), however, it does not match the means and goals, expressed in HCI and usability literature (see also Courage & Baxter, 2005). Rather, it resembles a 'reference interview' (as exemplified by Ross, Nilsen, & Dewdney, 2002), which has the sole purpose of "[f]inding out what they really want to know" (2002, p. 67). Establishing a good interpersonal contact, asking open-ended questions (especially 'why?' questions), paraphrasing and paying attention to the way the participant expresses themselves. In addition, the knowledge backgrounds of the participants can be looked up by searching for their contribution to the field on the Internet. Still it is anticipated that the interviews demand a high level of improvisation.

questions are crucial in a pilot study; they enable the participants “to highlight the issues that they find most relevant” (1998, p. 65). Still, fixed-response Likert scales are implemented to measure the degree of comfort, effectiveness and efficiency, which will help to judge how much weight the qualitative comments have.

After the meeting, the recording data is compiled into a Morae project for analysis. Here, Jordan's differentiation between *minor*, *major*, *fatal*, and *catastrophic errors* (1998, p. 20f.) is applied, alongside with the given analysis tools and frameworks of the Morae manager software, especially the 'markers', a coding function.

Results

In total, 25 potential participants have been identified and contacted all over Denmark; nine of these responded, eight of them positively. This gives a positive response rate of 32%. However, time constraints on both sides, as well as logistics made it challenging to schedule meetings. Most respondents noted that they had a very tight schedule at the moment and might have time only after this study's assignment deadline. This was additionally drastic, because their answers came with a large delay (some interested experts took over two weeks to respond; by that date, it was not possible to arrange a meeting any more, or even have time to analyze their data). In the end, only three meetings could be arranged and executed.

Participant 1 (P1)

P1 is a history student at a university in Copenhagen and part of a departmental project for improving the faculty's learning environment. According to P1, this job does not just include dealing with physical space, but much more with services, such as academic literacy, alternative content delivery, and workshops. In short, they are an innovation hub, with focus on student involvement. From this situation, the natural search task “*How do I get students engaged?*” was constructed, because P1 had never truly researched the topic and during the briefing interview, the question of student engagement arose frequently. The test took place in the project office in the respective department.

The briefing and the search task interview took around half an hour, as expected, without complications. However, the test duration surpassed the anticipated frame; nearly one hour was taken for the task and the survey. P1 made a positive impression at the debriefing, suggesting that he was satisfied with the search results. The survey results supported this satisfaction, however, upon assessing the recorded data, it became apparent that the test time was so long because P1 quickly got lost on the webpage and took a lot of the test time attempting to navigate back to the original page – the Zotero group. P1 even decided to call in the investigator in, to help him out (30:54) – in a real situation, P1 would have possibly given up –, a fatal error.

Participant 2 (P2)

P2 is a director of a higher education institution in Copenhagen. The author identified him due to his contribution to a project on primary schools, but he has contributed to far more projects, as he reveals. In a 20-minute briefing, P2 explained the infrastructure changes at his institution, the concepts, and some feedback. Just as with P1, the school was seeking contribution and evaluation by students and staff, but it is differently organized. The evaluation procedures were in total not as thorough and methodic. P2 expressed a great interest in a diachronic approach to learning space development, contrasting the past learning culture with the emerging contemporary. The latter was to be 're-ritualized' and 're-materialized', in order to create a new individual and informal learning culture. At the moment, this 'culture' was not yet 'stable', or not yet accepted as a usual practice. Therefore, the following

natural search task was formulated: *"How can we re-ritualize and re-materialize learning spaces to boost work motivation?"*

During the test, which took place at the office desk of P2, the participant had to consult the instructor a few minutes after the start, because of an unstable Internet connection with EDUROAM. After a hotfix through the instructor, the test continued for about 15 minutes. P2 made a very positive impression in the debriefing. Although he called himself not a researcher (as he normally has other employees and partners delivering research and information for him), but a decision maker, P2 mentioned that he found many interesting documents. He said that a (visual) guide on how to use this database could be a great help, including tips for the free text search. The classification structure appeared very intuitive and helpful to P2.

Upon importing the recorded file into the Morae project, the Morae manager issued an error message⁸. The recording appears to be corrupted and cannot be used for assessment. Hence, sadly, only the debriefing input is of value for discussion.

Participant 3 and 4 (P3 and P4)

P3 is an architect, involving mostly in designing classrooms and primary school buildings. P4 is his daughter, a teacher and anthropologist, who has been traveling and working around the world. Both own an architectural consultancy and embody a great amount of experience and knowledge in the field. The investigator visited them at their home, and from many perspectives, this turned out to be a very unpredictable, but positive meeting.

Their interest, openness, and hospitality has been beyond the investigator's expectation, so that it became challenging to conduct the field study in its scientific wholeness; for instance did they not see a need for me to be absent throughout the tests. Moreover, they were so enthusiastic about the topic that P4 began searching on her tablet computer before the search tasks had even been defined. In total, their general working style turned out to be a very collaborative one; they would normally search on their own computers in the same room, while continuously communicating, sharing and interacting. Therefore, it was agreed upon a collaborative test run, where both examine the database at the same time. Their persistent communication had the positive side effect that they articulated their thoughts and opinions, so that they could be recorded.

Another unanticipated detail was their speed in coming up with a natural search task. In split seconds, P4 formulated her research interest: *"Problem based learning spaces"* (in a pedagogical perspective); and P3 amended *"in the US, Germany, Norway"*, while also mentioning, that he, as an architect, would likely rather start by searching for *"open spaces"*. They articulated in advance that both of them had very different search styles, but they generally felt that the combination of the two yielded the most productive results.

During the test run, just as with P1, they once required the investigator to show them how to navigate back to the original web page. In the debriefing, they explained that they were very confused by the purely text-based architecture. Planners and architects, in their view, are visual thinkers, who deliberately search for visuals and pictures, before assessing a text, or even only the abstract. P4 illustrated her usual document assessment habitus as "flicking-through" document pages on her tablet, principally only paying attention to pictures; only if any of these caught her attention would she consult the table of contents, and even later, the abstract. Hitherto, the database contains no pictures whatsoever, and the few "kinesthetic" structures (such as the document icons) were not very satisfactory or intuitive for them. In their conclusion, Zotero, in its current state, presents an expert tool for information specialists,

⁸ The error message reports that the „recording cannot be opened (internal hardware configuration file error).“ Even after restarting, the error persists, and no hotfix documentation can be found in the support forums of TechSmith. A screenshot of the error message is attached in the appendix; the recording of P2 is stored alongside with the Morae project.

who are used to relational databases and intellectual data enrichment – a short manual, or even a video about how the search tools would could already make up for that.

Upon the end of the debriefing and vocal feedback, they requested another test run in the database, since they wanted to try it out now that they had some learning experience. After this second test run, they requested from the investigator to demonstrate them the *intended* and best use of the database. It became clear that the interoperability of the search methods was not intuitive for them, so that they would only choose a very low category (2) and browse the list of results, instead of further faceting it, which would be required for more satisfactory results in such a precise search task as theirs. In the end, they were very fond of the idea of collaboration that the Zotero group holds, in order to enrich the interaction in their professional network, and to draw inspiration from the architectural advances around the globe. In total, the meeting lasted more than three hours.

Data analysis

Despite only two samples of recorded data, and quite a strong difference in background of the participants and their environment, the Morae recordings show a good number of resemblances, when comparing the coding ('markers' in Morae). Both times, the first action of the participants was to navigate *away* from the database. Seemingly, they are attracted by the contrast heavy menu at the top (see Figure 1); this menu is not related to the Zotero group itself, though; it only concerns the Zotero product. For a considerable long time, they spend their time on the Zotero website, outside of the Zotero group. After a couple of minutes of attempting to return to the original page, they give up and consult the instructor, which suggests a fatal error.

For the time that P1, P3 and P4 spend on the intended Zotero group page, they apply similar search styles, prioritizing the free text search options (1). Both oversee the link to the "group library" page at the bottom of the start screen, where the three search interfaces are located; they only get there by browsing and selecting the most interesting item from the "Recently Added Items" list on the start screen. P1, is the only one who uses (1) inside the Zotero group, where the database indexes are actually taken into account for retrieval. However, he either retrieves far too many, or zero results. This occurs, partially because of misspellings (despite that P1 often uses Google Translate) and Zotero does not feature an auto-correct or recommendation functions; partially because he is simultaneously in the semantically wrong category (2)⁹. Only in the very end of the test, P1 discovers the semantically correct categories that best suit his research question (he chooses the categories 5.5, 5.5.1, partially 7). He expresses in the survey and in the debriefing that these documents were indeed very interesting, also because of their international diversity.

⁹ To outline this major or even fatal error, when selecting a certain category (2) or tags (3), the free text search (1) can be used to search further in the chosen categories and tags. P1 did not know this search function relation and therefore expressed confusion about why the simple query "student engagement" did not yield any results in the category "5.1 organizational structure".

The screenshot shows the Zotero website interface for a group named 'Learning Spaces - Lernorte'. The page features a navigation bar with options like Home, Groups, People, Documentation, Forums, and Get Involved. A search bar is present in the top right. The main content area is titled 'Learning Spaces - Lernorte' and displays a list of 'Recently Added Items'. Each item includes a title, the user who added it (sv034), and the date and time it was updated. The items listed are:

Title	Added By	Updated On
Lernarrangements in Bibliotheken - Support für informelles Lernen	sv034	May 26, 2014 9:11:02 PM
Räumliche und gestalterische Anforderungen an Bibliotheken als Lehr- und Lernort zur Förderung von Informationskompetenz	sv034	May 26, 2014 9:07:11 PM
Informationskompetenz und neue Lerninfrastrukturen in der Hochschulbibliothek	sv034	May 26, 2014 9:00:47 PM
Designing Good Library Space to Promote Information Literacy	sv034	May 26, 2014 8:48:31 PM
Students' Perspectives towards Learning Centers	sv034	May 26, 2014 8:33:48 PM
New Learning Spaces at German Universities: A Survey Initiative of the Task Force Learning Spaces of DINI	sv034	May 26, 2014 7:58:30 PM
Lernwelt Universität: Entwicklung	sv034	May 26, 2014 7:43:47 PM

On the right side, there is a profile for 'HOCHSCHULE DER MEDIEN' with a description: 'Wandlungen im Bildungsbereich sensibilisieren Schulen, Universitäten, Weiterbildungseinrichtungen, Bibliotheken und Museen zur Neugestaltung von Lernumgebungen. Mit einem Schwerpunkt auf Bibliotheken, Erwachsenenbildung und Trends, wie Kooperation und räumliche Integration von Bildungszentren, soll diese Literatursammlung Entwicklungen im Bereich physischer Lernorte auf nationaler und internationaler Ebene aufzeigen und Hilfestellung zur Konzeption, Gestaltung, Organisation und Umsetzung von attraktiven Lernumgebungen bieten. Initiiert unter Prof. Dr. Richard Stangs Forschungsschwerpunkt "Lernwelten" an der Hochschule der Medien (HdM) Stuttgart, wird sie von Stefan Volkmann (Bibliotheks- und Informationsmanagement, HdM Stuttgart) gepflegt.'

Figure 1: The screen setup after the Morae recording is started

P3 & P4 discovered the category system (2) and the tags (3), as well, but did either not use them, failed understand their utility, or experienced fatal errors, connected to the Zotero website¹⁰. (P2's data cannot be analyzed, but he mentioned an intuitiveness of (2), suggesting that he did not run into usability issues there.) It makes the impression P3 and P4 simply expected a very different way of collection display. When browsing through other Zotero groups, they were oftentimes looking at the profiles of other people¹¹, who added an item that was somewhat in their interest. Especially P3 frequently expressed his expectation for visual materials. When evaluating information items, P3 and P4 were spending the most time over online documents with pictures, to discuss them. As Zotero is a purely bibliographic and text based system, they did not find what they were looking for and this reflects in the low satisfaction rankings in the survey.

The tag system (3) was least used, and it did not make the impression that either participant understood what it was good for. In general, the interoperability of the three search methods did not seem to become very apparent. Only during the last quarter of his test did P1 seem to realize that selecting a lower category (2) broadens the scope of the results, and that (1) can be used to search inside the selected category. However, again, search attempts with (1) yielded too few results, because of only a partial semantic overlap. In general, P1 seemed to have grasped the concept behind the categories (2), as he correctly chooses category 5.5.1 and 7 for his search task. Still, in his survey, although mentioning a simple browsability of the collection, he insists on (1) as his preferred search method and even requests "advanced search" functions. For P3 & P4, the database was not worth consulting again – as far as their survey goes (they expressed themselves much more positively during the debriefing). For them, the navigation has to be enhanced, in total.

¹⁰ With regards to the latter reason, P4 was in the habit of opening links in new tabs, but this method did not work when applied for the tags – only an empty page opened in the new tab.

¹¹ Considering their professional interest in their 'network', as well as their input in the survey and during the debriefing, it becomes clear that they had a strong interest in finding information, but also the people that produce and share them.

Discussion

Given the very few (successful) samples of this pilot study, there is too little data available to address the research question and the hypotheses in a reliable and valid manner. Even within the two samples available, the data analysis uncovers far more usability issues than the research question is addressing, however.

Differentiating between the Zotero website and the 'Zotero group'

Besides the search functions, the Zotero website itself is a major distractor. Too often did the participants mistake the menu on the top of the Zotero webpages as integral functions of the database. (Moreover, for P1, upon attempting to return to the original page (using the 'return' button), a website error occurs (11:32)). All together catastrophic errors, which required help from the investigator in both recordings. In a real-life situation, none of the participants would have managed to find and consult the database, and would quickly have given up and resorted to their usual means of retrieving information (which was mostly Google products, but also Academia.com (P4)).

The fact that the participants did not know Zotero likely led to the reasoning that the tested database itself was called "Zotero" (especially P3 and P4), which would explain why they selected the main menu buttons when they were trying to start from scratch, or find help on how to use the database. Consequently, the participants may not know what a "Zotero group" is, and how it is different from the website Zotero (which is only providing the hosting). Moreover, especially P3 & P4 did not seem to recognize the search functions as the major means to browse and search in the collection – in their own terms: they were looking for pictures. Their "kinesthetic" way of thinking makes them look for visual rather than text, and therefore more or less 'blind' for the intellectual and bibliographic construction of Zotero. Apparently, according to P3 and their survey input, the debriefing discussion even led to the perspective that Zotero might be unsuitable for architects and policy makers, who allegedly prioritize visual components, especially for "initial search". In the lens of this argument, Zotero, as a tool, may itself carry a structural bias in being designed by and for information specialists and researchers. P3 and P4 imply that the design does not appeal to a non-scientific community. Their critique extends also to the documents themselves, being structured 'as if they were 10 or 15 years old' (P3): plain and stiff text with no visual enhancements to draw interest and attention to. In short: how can the goal of the database (connecting the various research hubs around the globe to advance the movement) be reached, if it is only designed for an expert community? Indeed the same question has to be asked when assessing the three search methods.

P3 and P4's critique is justified; in retrospective, the investigator has imposed a librarian's perspective on what a satisfactory search result may look like: a collocation search – finding every relevant document on a certain topic. However, this may not be the criteria that the population judge the success of their search activity, as the second sample illustrates. The participants' positive attitude after test runs full of fatal and major errors suggests that they may only look for a few or a single document to inspire them. The investigator agrees that this pre-imposed assumption is embedded in the structure of the collection, and in Zotero itself.

Assessing the three search methods for user satisfaction

As the participants express, they wish to use the free text search (1) as their primary search option. This appears to confirm the first alternative hypothesis, but it does not mean that (1) performs satisfactory – quite the contrary. Besides the condition that the user has to be located in the 'group library' for the search function to address the collection index (otherwise, the *Zotero website* is searched in), a number of other reasons add up to the status quo of (1). Opposing to search engines, like Google, yielding thousands of search results in split seconds, this database only includes approximately 1500 information items hitherto. First, the users will be disappointed, if they expect a similarly fast performance and results rate. Secondly, the search bar of (1) has the relatively hidden option to search either in

"Title, Creator, Year" or in "Full Text". If this detail is not noticed and not understood, search results may not be as satisfactory. P1 realizes it only after 14 minutes, but retrieves no better results because of spelling mistakes in his queries. Missing spell correction therefore also adds up to dissatisfaction. Furthermore, as a collection of very heterogeneous sources, a controlled vocabulary, as used in (2) and (3) is necessary. Because the various national research hubs attribute different names to similar concepts (also using different languages), it is impossible to achieve a high recall and precision only searching with the typical one or two search terms (Ingwersen, 2005, p. 126) – a behavior, visible in both samples. On the contrary, in the debriefing, P1 pointed out that a multitude of information items are simply called "learning spaces" (or similar generic titles). This makes it not just difficult to assess the content of an item without thoroughly reading the entire bibliographic entry (which supports P3 and P4's request for more visual elements); it also complicates the precision of (1).

For both natural search task of P1 and P3/P4, (3) would have been the most ideal, regarding recall and precision, because of the possibility to combine the controlled categories of (2). Facetted correctly, the result list contains all relevant information items, cross-linguistically. Henceforth, more attention needs to be given to designing the tag system (3) in a more prominent and intuitive manner, so that it appears as an easy to use alternative to (1). (As P1, P2, and P3 agree, a simple introduction to the website and the search methods would already make a big difference.) Still, as the general interface usability appears to impose more orientation struggles, there was no particular observations that supported that the descriptor formulations of (2) and (3) have a particular impact on the usability (P2 even addresses them as "intuitive". Therefore, the second alternative hypothesis cannot be confirmed.

Conclusion

Despite only two useful samples available for analysis, the pilot study has been quite successful, regarding the anticipated and discovered usability issues of the assessed Zotero group. As the research question of this paper states, only the *search functions* of the Zotero group "Learning spaces – Lernorte" was to be evaluated. However, the recordings quickly illustrated that a lack of experience with Zotero imposes holistic navigation issues, followed by search retrieval issues. These turned into fatal and catastrophic errors, oftentimes. Considering that the author has no direct influence on the interface design of the website (only via the bibliographic data can the usability be manipulated directly, when using a Zotero group), and the fact that Zotero is an open-source initiative, the interface design may be quite inert for change. However, the findings suggests that fundamental interface changes have to be applied in order to ease the access for users without prior knowledge of Zotero.

Applying a visual dimension to Zotero, as P3 and P4 request, will be rather challenging, because it would require a fundamental change of the Zotero software structure. More realistic is Heimburger's approach of using the RSS and API of Zotero to 'outsource' a part of the collection display and navigation on another website. This could be a more user-friendly approach, because it gives more flexibility in designing an individual solution to usability issues. Here, also the proposed user manual, or introduction videos could be embedded, and there would be space to relocate and redesign the search function interfaces.

How exact changes should look like in order to boost user satisfaction cannot be reliably deduced from only two samples and four participants. Indeed, the three search methods seem to be used in different degrees and with different satisfaction, as the first alternative hypothesis suggests. However, their ideal 'best-practice' use and the way the participants used them are not identical. In the following, the three search methods in Zotero are shortly assessed in terms of the user interaction and satisfaction, and their objective value.

1. The free text search option (1) was most preferred by the participants P1, P3 and P4. However, it frequently yielded dissatisfactory results, reflecting in low effectiveness and efficiency ratings. From the discussion, it seems as if (1) was entirely unsuitable to be applied in this database (except for known information searches, like searching for works of a specific author).

2. The category structure (2) above the tag cloud was only marginally explored, but it seemed as if it helped P1 to retrieve the most relevant of his findings, and gave him an initial understanding of the interoperability of the three search methods. Also taking into account P2's positive opinion on (2), the browsing seems to be easier and more comprehensible than in (3); it appears most suitable for wide research questions and unexperienced users.
3. In all cases, from an objective perspective, the combinatory feature of the tag cloud (3) would have yielded the best results to fulfil the natural search tasks. This search method was never seen to be applied by the participants in its full functional spectrum, and was possibly never even understood fundamentally. It would be most fruitful to apply Heimbürger's strategy with this search method, so that the website that the collection display is embedded in, shows not just an indifferent collection of tiles (representing the categories of (2)), but allows dynamic combination of categories, and thereby faceting of long result lists.

Perspectives for further research

The observational findings and the concluded practical suggestions of this pilot study converge with further usability research. Future studies should examine how the three search methods are to be designed and positioned, for users to *intuitively identify the one that suits their information needs best*, without being distracted from the Zotero website itself. The overall goal is achieving a larger, more reliable sample, and discovering best practice design solutions that further usability.

- *How can the sample be increased?*
 - o *Longer time frame* – As the return rate showed, the issue with recruiting participants is not getting their interest, but scheduling a meeting with them. Even after two weeks, some contacts responded positively to the initial enquiry, and some of them were proposing a meeting only in another few weeks. In order to account for this, a total timeframe of about two months (from first contact to final analysis) is mandatory to make use of the local participant supply, and to give enough time to plan and execute travels.
 - o *Extending international reach* – as the method discussion highlights, the research design is unsuitable for virtual tests across borders. In order to account for an international population, all hardware requirements have to be eliminated from the test design. One could reduce the experiment to usability surveys or diaries, where the natural search tasks are agreed on during a preliminary online interview. The depth of data will only be slightly more detailed than P2, though. Alternatively, virtual screen capturing could be accomplished using the screen capture features of internet telephony technologies, like Google Hangout; the investigator's screen can then again be recorded, using Morae. Using such technology combinations, a 'home-visit' is no longer required and participants all over the world become prospective. The raw data may be less detailed and more difficult to analyze, but in total, a larger sample can be assessed and regional biases eliminated.
- *Is there an alternative to Morae?* – Despite a certain mobility, TechSmith's Morae does not qualify for actual field studies; at best, it is a usability laboratory on a laptop. Throughout the tests, far too many distractors disturbed and conditioned the behavior of the participants. Ideally, the experiments should be carried out on the participants' own devices; otherwise, no valid realistic field is established. Simply using a regular camera, placed beside the participant's work desk could be a fair trade-off¹².

¹² Concretely, it should be at least two cameras – one monitoring the screen, and one recording the participant. The synchronization can still be handled using Morae, where the cameras are just used instead of the screen recording input. The only distractors in the natural work field of the participant will then be the cameras, tripods, and the connection cables to a laptop, where the test is recorded.

- *Was the start page of the test well chosen?* – When searching for the database on the web, a multitude of ballast entries linking to various of the database's information items appear as well, suggesting that at least a large part of the collection is indexed in Google. From this, one could assume that people would run into the database because of a link to a bibliographic entry in Google. Then, the first page they see already features a document they perceive as relevant, as well as the three search interfaces.
To embed this component into the research design, the participant has to be asked to do an initial search in Google on his search topic (under supervision of the instructor), right after the recording starts. The query is amended by "learning space zotero". As long as the query scope is broad, there should appear a relevant deep link into the database collection; else, this step has to be repeated.
- *How can one account for the heterogeneous backgrounds of the participants?* – At the moment, the heterogeneous project target group – the population – imposes uncertainties, because the backgrounds of the participants influence their information behavior, their experience, and their expectations (see also Lazar, 2010, p. 9). For instance, P3 and P4, an architect and an anthropologist, had tremendous problems navigating in the very text based interface. In order to account for this suggested 'visual fixation', the occupation of the participants have to be taken into consideration. Without introducing artificial within-group search tasks, it simply has to be made sure during the participant scanning that all stakeholder groups are to a similar degree accounted for. Conversely, a within-group design may be necessary to further address *efficiency* and *effectiveness* of the database, in comparison to a participant's usual research methods.
- *Do participants perform more satisfied when being introduced to Zotero beforehand (e.g. through specific manual materials)?* – There was substantial, even if only few evidence that the participants prefer short introduction materials. The design of such materials should hence be taken into account during further studies, as a treatment, alongside with control groups.

References

- Ancient World Open Bibliographies. (2013, November 10). *Zotero*. Retrieved May 19, 2013, from https://www.zotero.org/groups/ancient_world_open_bibliographies
- Collaborative Bibliography. (2014). *International Society for First World War Studies*. Retrieved May 19, 2014, from <http://www.firstworldwarstudies.org/bibliography.php>
- Courage, C., & Baxter, K. (2005). *Understanding your users: a practical guide to user requirements methods, tools, and techniques*. San Francisco, CA: Morgan Kaufmann Publishers.
- First World War Studies Bibliography. (2014). *Zotero*. Retrieved May 19, 2014, from <https://www.zotero.org/groups/55813>
- Heimbürger, F. (2013a, March 25). *Nouvelles pratiques collaboratives pour la recherche internationale. L'exemple de la bibliographie de la "First World War Studies Society."* Conference presentation presented at the De la diversité des humanités numériques: une exploration des pratiques, Paris. Retrieved from http://www.canal-u.tv/video/ehess/3_de_la_diversite_des_humanites_numeriques_une_exploration_des_pratiques.12910
- Heimbürger, F. (2013b, June 13). *Veiller, communiquer, gérer une bibliographie collaborative avec Zotero pour la First World War Society*. Presented at the Digital History, Toulouse. Retrieved from http://www.canal-u.tv/video/universite_toulouse_ii_le_mirail/veiller_communiquer_gerer_une_bibliographie_collaborative_avec_zotero_pour_la_first_world_war_society_franziska_heimburger.13975
- Heimbürger, F. (2013c, September 4). *FWW Studies Fachbibliographie - 9.4.2013: Präsentation des kollaborativen Bibliographie*. Prezi presentation. Retrieved from <http://prezi.com/aie366skczya/fww-studies-fachbibliographie-942013/>
- Heimbürger, F. (2014, May 18). First World War Zotero bibliography project. Direct correspondence with the author.

Hiltmann, T. (2014, January 24). Collaborative bibliography on heralds on Zotero. *History of Heraldry*.

Weblog. Retrieved from <http://heralds.hypotheses.org/46>

Horne, J. (Ed.). (2010). *A companion to World War I*. Chichester, U.K. ; Malden, MA: Wiley-Blackwell.

Ingwersen, P. (2005). *The turn: integration of information seeking and retrieval in context*. Dordrecht:

Springer. Retrieved from <http://ep.fjernadgang.kb.dk/login?url=http://dx.doi.org/10.1007/1-4020-3851-8>

Jordan, P. W. (1998). *An introduction to usability*. London ; Bristol, Pa: Taylor & Francis.

Lauterbach, D., Sun, K., Totz, K., & Wang, W. (2009). *Zotero Usability Testing: Executive Summary* (p. 24).

University of Michigan. Retrieved from <http://www.debralautebach.com/projects/zotero.html>

Lazar, J. (2010). *Research methods in human-computer interaction*. Chichester, West Sussex, U.K: Wiley.

Learning Spaces - Lernorte. (2014, May 17). *Zotero*. Retrieved May 19, 2014, from

https://www.zotero.org/groups/learning_spaces_-_lernorte

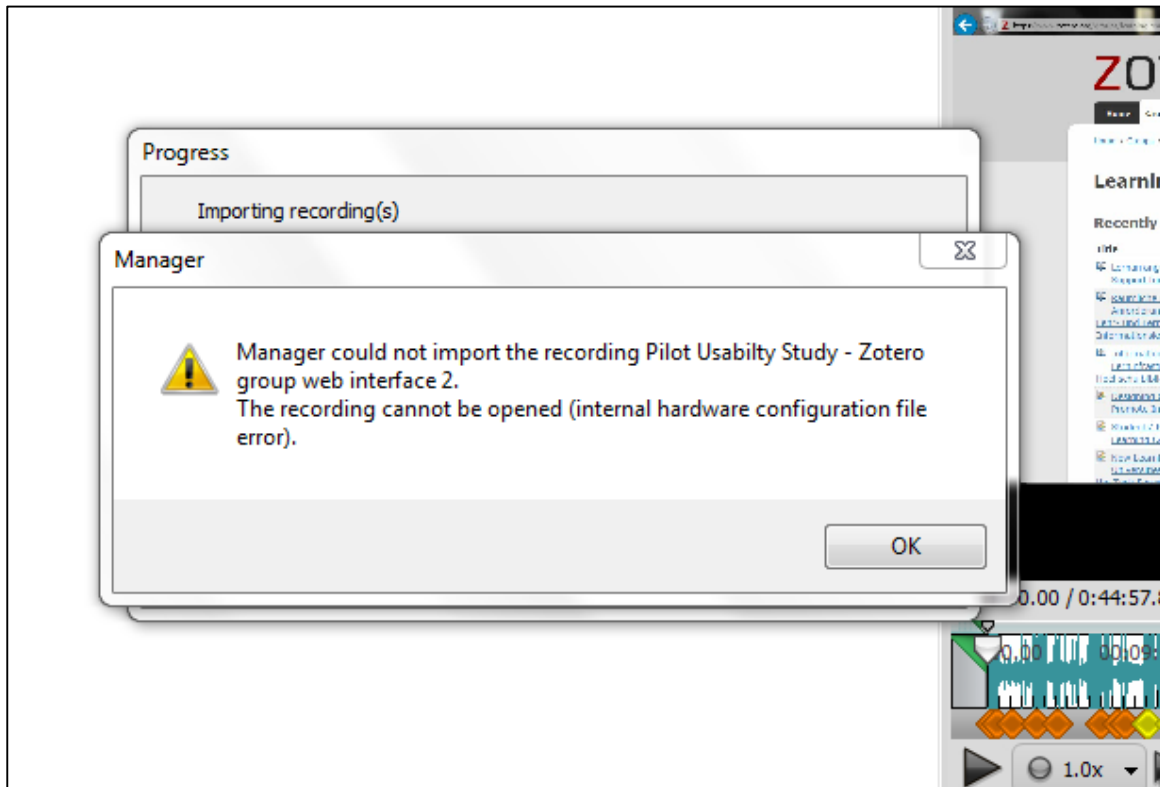
Ross, C. S., Nilsen, K., & Dewdney, P. (2002). *Conducting the reference interview: a how-to-do-it manual for librarians*. London: Facet.

TechSmith Corporation. (n.d.). Intermountain Healthcare Conducts Large Scale Usability Testing in the Field.

TechSmith. Morae Case Study. Retrieved May 26, 2014, from <http://www.techsmith.com/morae-casestudy-intermountain.html>

Appendix

Morae manager error message for recording of P2



Morae post-recording survey

As a last step, please fill out this usability survey, in order for us to evaluate, how satisfied you are with the task results and for any comments that you have.

1. Have you seen and used this Zotero group before?

If so, can you remember what you were looking for and if you were successful?

- No
 Yes

2. During the test, did you feel as comfortable as you normally feel when working alone?

Was there anything distracting in the environment?

- 1 2 3 4 5 6 7 8 9 10
I felt very uncomfortable I felt very comfortable

3. Are you satisfied with the results you got from the search task?

Why?

- 1 2 3 4 5 6 7 8 9 10
Very unsatisfied Very satisfied

4. Were there any particular findings you were positively surprised about?

5. Was there any topic you expected to find more information on?

6. Considering the time and effort you put into the search, was it efficient to use this database, compared to your usual ways of researching?

1 2 3 4 5 6 7 8 9 10
Very inefficient Very efficient

7. Which of the three search methods did you use the most?
If you used a combination, please elaborate!

- Free text search through the search bar in the top right corner
- Browsing the collection through the folder structure on the left
- Browsing the collection using the tag cloud at the bottom left

8. Why did you prefer this search method over the others?
What difference did it make?

9. Was there anything particularly distracting about the website interface, in general?

10. Would you recommend this database to your colleagues?

1 2 3 4 5 6 7 8 9 10
Very unlikely Very likely

11. Will you consult this database in the future?

1 2 3 4 5 6 7 8 9 10
Very unlikely Very likely

12. Any comments you would like to add?
You may call in the investigator again.

Statement of informed consent

Statement of Informed Consent

Stefan Volkmann
 ERASMUS exchange student at
 Copenhagen University
mvi657@alumni.ku.dk
 +45 29417772

Purpose:

You are volunteering to support the *user-friendly development of the forthcoming largest international research database on physical learning spaces*. The project is initiated and funded by Media University Stuttgart, Germany and under supervision of Prof. Dr. Richard Stang. Stefan Volkmann, who is also conducting the tests with you, is the major project coordinator and database curator. He will brief and guide you through the usability study, and will give you an in-depth introduction into the project and the collection afterwards, if you request so.

As an expert/practitioner in the field, you are our major target group. Our goal is to design and structure the database to benefit your research needs and behaviors. The test does not intend to measure your individual performance.

Procedure:

1. Through an interview, we **agree on a search task** that is *in your personal/professional interest*
2. The **search task is carried out by you on the laptop** made available by to you. Your search behavior *will be recorded* through TechSmith's *Morae* usability software. It includes *on-screen capture, webcam video, and voice recording*. The *investigator will be absent*, so that the environment will resemble your usual workspace as much as possible. Feel free to use your usual research sources as well, if the database does not yield satisfactory results.
3. You may **break off the task whenever you feel like it**. Upon stopping the recording, you will be asked to fill out a **survey** that asks for your satisfaction with the search task and the database.
4. You may invite the investigator into the room again. He will **debrief** the usability study and you may *share comments or ask questions* that you have regarding the test or the database.
5. Your data will be **analyzed until the 10th of June**. Thereafter, the report is assessed by the university and **your data will be deleted by the end of June**.

Confidentiality:

Your data will remain anonymous at all times and it will only be used to assess the usability of the database. All raw data will be deleted by the end of June.

You may withdraw from the test at any point

If you agree with these terms, please *indicate your acceptance by signing below*:

Signature: _____

Printed name: _____

Date: _____