Digital Libraries – Literature Review.

December 15 2006 – submitted for University of Wolverhampton MSC module.

By Stacey Greenaway

Contents

1. Section A –Digital Libraries

1.1 Introduction.

Digital libraries are relatively new, evolving increasingly with the phenomenal success of the internet. This has opened up the idea that digital collections can be made available to a wide variety of users, over an international platform. With this innovation comes a limitless set of problems to be overcome through research. In this review, the broad outline of the types of problems facing digital library development with be discussed resulting in particular focus on usability. This vast area of research has been narrowed down to techniques for evaluating the usability of digital libraries and the shortcomings of these standard techniques that work well on the internet platform, but not for the digital library domain. Ultimately, these shortcomings are investigated through discussion of research into proposed frameworks for evaluating multiple digital libraries attribute by attribute in order to begin to create some standards for digital library development.

1.2 Broad Topic – General

The web based nature of digital libraries means research into internet/WWW based topics are relevant to digital Libraries, e.g. accessibility, usability, information retrieval, automation, xml, metadata; however, whilst this research provides a grounding for digital libraries, their diversity brings about a sub set of research areas that can only be relevant to their domain. In addition, there are areas of research relevant to the information science aspect of digital libraries, where traditional library ideals are being transferred onto a technical platform for international audiences. One of the greatest problems for digital library development is fulfilling the requirements of Computer Scientists, Information Scientists and the myriad of users that will access the collections.

There are countless different types of digital library e.g. music, video as well as academic collections, which will have their own set of requirements for research. There are some general areas that are relevant to all libraries regardless of genre. Classification of digital libraries is one area of research that has implications for other research areas such as usability and information retrieval, in that it allows for digital libraries to be compared to each other, this is helpful in searching collections but also in evaluating their usability. One key way of classifying digital libraries is to produce metadata, which is used to index all of the content in the library; it provides information about each item in the form of keywords and abstract descriptions that can be searched. This metadata can be produced manually, but research into automating its creation is being conducted. By automating the digital library, this metadata can easily be entered by a "librarian" and a new item added to the library. Further research in this area includes algorithm designs to automatically capture the metadata out of scientific journals and conference papers in order to categorise them. More research areas include using xml to do this, which crosses over into research for information retrieval. This is a well established research area for the internet, researchers are currently experimenting with ways of transferring some of the new developments for web searches such as clustering, and topic based searching to retrieve information from digital libraries more quickly and accurately inline with users information needs.

Research areas which cross over between computer and information science are Obtaining Information, Preservation and Quality of Service. Obtaining Information can be defined as, who provides the content for the library? Problems can arise if institutions are not willing to freely share their resources. Depending on the type of

library, information may need to be digitised, or converted to more recent technologies. There may also be issues of trust, whereby owners of the collections may not fully understand internet technology or dislike the lack of control they have over access to their collections. Preservation of information has been a concern for traditional librarians since their creation and while many techniques exist, more research is needed into ensuring the data in the digital library will be accessible in the future when technologies change. When designing a digital library the system needs to be able to adapt to changing technologies. In the past inventors/scientists/authors/philosophers kept notebooks that have been preserved and contain background to some of their major works and theories. Nowadays these notebooks exist not physically on paper but as attachments in email or in emails themselves, or maybe as blogs. The question of how all this information is captured and preserved for future generations to be able to read is an important research area of digital library development. Quality of Service exists as a research area generally for distribution of digital content across the internet and so is of importance in digital library development. Ensuring the quality of information inside the library includes checking the validity of the research when creating a digital library of academic/scientific research papers. File quality is important in video and music libraries. For images, the quality of the image includes optimisation but also that it is not skewed, not to dark cropped wrong etc. This crosses over into usability and accessibility principles in that the content must be of good quality and easily readable by the majority of users.

A research area relevant to information science is in the area of Intellectual property rights and Copyright laws. There is contrasting opinion toward restricting open access to information versus freely sharing information, which the internet was created for. Research areas include how intellectual property rights and copyright laws are adhered to whilst still being able to freely share information. How can the information be protected against misuse e.g. plagiarism, misquoting? This area is of particular importance with music and video libraries where it is vital that permission is granted by the artist/publisher and in most cases, the correct fees paid.

The above discussion draws on the opinions of McCray and Gallagher (2001).

1.3 Broad Topic: HCI, Usability and Accessibility

McCray and Gallagher (2001) give an overview of issues concerning HCI, Usability and Accessibility of digital libraries. The system must be accessible to the vast majority of users regardless of disability, language or cultural differences. The information should be easy to find using keyword searching. The interface should be intuitive with the keyword search easy to find as well as the ability to browse topics. Information should be optimised so that the retrieval of information is quick. Good indexing is vital so that the search terms can be well matched in results and results found quickly. Further research areas exists within the usability area for accessing digital library collections on small screen hand held devices, in measuring the users experience when interacting with a digital library which is linked to research into emotive design, and in cross-cultural usability.

2. Section B - Usability Evaluation of Digital Libraries.

2.1 Usability Problems

Several usability studies of digital libraries exist namely, Alexandria Digital Library, Greenstone Digital library, Perseus Digital library and NCSTRL (Networked Computer Science Technical Reference Library). The latter was evaluated using usability inspection methods in a case study by Hartson et al. (2004), the analysis of the evaluation raised numerous problems found in evaluating digital libraries. A lack of research was highlighted in techniques to effectively evaluate digital libraries along with the issue that they, as usability specialists, do not have extensive knowledge and experience with digital libraries and can only evaluate using general guidelines established for systems; highlighting the research requirement for a usability evaluation technique specific to digital library evaluation. I can agree that although research exists into usability evaluation for digital libraries, it is not extensive which is surprising as usability is of vital importance in development of any system.

Saracevic and Covi (2000) make an interesting argument that defines a usability problem for digital libraries, traditional libraries are all organised in a similar way and if you can use one library you can use another regardless of geographical location (language barriers aside). This highlights the need for standards to be developed in the design of digital libraries to gain "uniformity for access and use". They suggest that through the creation of an evaluation framework like their conceptual framework it may be possible to highlight common usability problems, in fixing them some standards will be established.

A major problem in developing any evaluation framework is in how to model user behaviour. This problem is highlighted by many researchers and will be discussed throughout later sections of this review. There has been research conducted that concentrates predominantly on understanding users of digital libraries. Adams et al. (2005) discuss users different information needs of different libraries and the "user's journey" in finding that information. They draw attention to the fact that "No sooner have users' needs been identified and supported than they change." Users' needs are very hard to model, causing a problem for usability design, there are different types of user with different information needs. Another problem is whether they enter what they want to search for into keywords, they may not have enough knowledge on the subject they are searching to enter correct search terms to retrieve the best results for their desired information need. Adams et al. also highlight the problem that different users interact with a system differently depending on their profession or subject interest area. In their research, they used users from academic and health care backgrounds and evaluated their comments on interactions with different libraries that highlighted the differences between users from different backgrounds and how hard it is to model these differences. This is a massive research area and beyond the scope of this literature review.

Blandford et al. (2001) discuss differences between users' interactions with the digital library when browsing and searching. A user's priority regardless of discipline is how easily they can interact with a system, how quickly they can receive results to their query and how relevant the results are to their query. This is the same for any information retrieval task. They also identify a problem when assessing the usability of multiple digital libraries, a good digital library means something different to different users depending upon their information need and background. How can a library dedicated to humanities research be compared to a medical science library? Equally, how can the users' needs for each be modelled when the designers and developers do not have sufficient knowledge of the content? Fuhr et al. (2001), Tsakanos et al. (2004) and Sandusky (2002) also discuss this problem.

2.2 Standard Techniques

One prominent aspect of research into usability evaluation of digital libraries is the idea of creating what Blandford (2004) describes as a "Toolbox of Techniques" that can be used to evaluate digital libraries at different stages of development. It is noted by Hartson (2004) that the majority of evaluation is done by analysts after the system is deployed and by analysts who have little experience with the information science attributes of Digital libraries and therefore the evaluation results do not highlight all the usability problems of the deeper system, only surface interface/interaction issues. As previously stated in Section 1, there is limited research in usability evaluation of digital libraries, as a result, one author and collaborators has been prominent in analysing usability evaluation methods, Blandford et al. (2004), they perform usability evaluations of multiple digital libraries using four different techniques and compare and contrast the results in order to try to find a set of techniques that can form a standard for digital library evaluation. The techniques compared were Heuristic Evaluation, Cognitive Walkthrough, Claims Analysis and CASSM (Concept-based Analysis of Surface and Structural Misfits). Early criticisms about this research are that firstly the research may be biased toward CASSM, as the technique was developed by two of the authors (Blandford and Connell), the authors have also worked previously with Claims Analysis to tailor it to the Digital library domain which may also provide a bias in opinion; secondly, different digital libraries were evaluated with each technique, for results to be accurately compared it would have been advisable to evaluate the same set of digital libraries with each technique.

The Authors found that it was impossible to evaluate every page of the Digital library using Heuristic Evaluation technique, due mainly to the size of the document repository but also due to the nature of the guidelines not being suited to the Digital library domain, for instance the guidelines do not allow for the process of forming queries to be evaluated, something which can cause a usability trap with in digital library systems. Whilst by focussing mainly on the interface and user interaction which Heuristic Evaluation and Cognitive Walkthrough do, can highlight general usability problems such as navigation or user feedback, it does not highlight deeper problems applicable only to the digital library. The authors highlight a usability problem common in Digital libraries in that an analyst can not know in what terminology a user thinks, they can not predict how much knowledge a user has in the subject when beginning an interaction with the digital library; the analyst may be familiar with specialist terminology used in the system that a non specialist user isn't familiar with, therefore usability problems may be missed. By creating user tasks for the Cognitive Walkthrough evaluation that were specific to this problem, the authors found that it did not highlight all the problems they were expecting and like Heuristic Evaluation returned only surface usability problems.

Claims Analysis is a technique where scenarios are established before conducting an evaluation, that similarly to user tasks in cognitive walkthrough, predict a user's interaction with the system. These scenarios produce positive or negative "claims" with negative claims being the usability problem. The cost of then fixing these problems is to try and remove the negative claims without affecting the amount of positive claims produced. The authors generated claims in three areas, user goals, user actions and system feedback. This method also allows for the creation of user profiles or "personas" which model the information needs and experience of the user. The authors found that by creating a persona for a user with the same level of expertise in the system as a system developer and a novice user with no experience of the system or little clear idea of their information need, a good range of usability problems could be highlighted. They state that in reality most users will sit somewhere in between these two extremes, I would argue that whilst this is a useful

tool, it is very difficult to model how users will perceive interactions with a system. There is discussion by the authors and also Fuhr (2001) and Saracevic (2000) that confirm the argument that most analysts come from either a Computer Science or an Information Science discipline and will perceive users interactions differently, they will also use different terminology and be interested in different aspects of the system, which will cause difficulties in modelling non specialist users and also affect the types of usability problems they identify. A problem highlighted by the authors with this evaluation technique is that to be fully effective it must be carried out as part of the software development lifecycle, where the in depth knowledge of the system developers can be utilised to ensure every section of the system is evaluated. This may sway the type of usability problem raised toward a technical aspect over information science problems, and whereas evaluation at this stage is vitally important to iron out technical bugs, for full evaluation of a digital library system it would appear that another technique would have to be used to highlight usability problems with an emphasis on information science.

The final evaluation method Blandford et al. analyse Cassm, considers the problems raised in analysis of Claims Analysis. Evaluations are performed with the specific aim of highlighting differences in the type of usability problems found when the evaluation is carried out with a team of developers, or in a stand-alone setting with non-technical users. As previously stated, this technique was created by two of the authors, who state that CASSM was created out of a "need to compliment" Cognitive Walkthrough and Claims Analysis when evaluating Digital library systems. CASSM is a technique of conceptualising a Digital library system, each concept is considered an entity and attributes defined for each entity. Relationships can be identified between different entities and attributes which allow for the usability problems highlighted (usability problems are named misfits in this method) to be analysed of either one system, where cause and effect can be measured of one usability problem on anther, or for comparisons to be drawn between multiple libraries that share usability problems within common attributes.

Blandford et al. conceptualise the Digital library system as two dimensional, User and System, they define five entities with in these dimensions, Users, Articles, Search Facilities, Feedback and Binders and the TOC service, they describe some attributes of these entities but do not provide a definitive list, the attributes will vary depending on the individual digital library. No reference is made as to how the entities are separated into the two dimensions, I would assume that the attributes of each entity are evaluated from either a user or a system perspective. The authors state a problem with the CASSM method in that it does not provide the detail in the usability problems it records that standard evaluation techniques like Heuristic Evaluation and Cognitive Walkthrough provide. This highlights a research area in using a conceptualising technique like CASSM to define attributes and then evaluate the attributes using a standard usability evaluation technique. The results of the authors' evaluations found that many common usability problems were highlighted by CASSM, which gives evidence that there is scope in the research area of creating a framework for evaluating Digital Libraries attribute by attribute in conjunction with a standard usability evaluation technique such as Heuristic Evaluation. This is a technique that has been proposed by other researchers Fuhr (2001), Saracevic (2000), Sandusky (2002) and Tsakonas (2004) among others, as a way of finding standard tools for evaluating multiple libraries and in highlighting usability problems that are common in all libraries, which can begin to form standards for Digital library development.

2.3 Usability Evaluation Attribute by Attribute

Through reading around the topic of Usability Evaluation attribute by attribute, two papers, Saracevic and Covi (2000) and Fuhr et al. (2001), stood out as preliminary research into conceptual frameworks for digital library evaluation. By comparing the two research papers to Blandford et al. (2004) it is clear where their ideas for the CASSM usability evaluation method originated.

Saracevic and Covi (2000) recognise through literature review that evaluation has been neglected in digital library research. They speculate that a reason for this is that evaluating a digital library is so complex it cannot be accomplished using standard usability evaluation techniques. They outline a requirement for a set of tools to evaluate digital libraries, much like Blandford (2004); Fuhr et al. (2001) also note this research aim. They also recognise the problem of how there are two distinct groups involved in the development and use of Digital libraries, whereas Blandford et al. (2004) identified these two groups as being technical specialists and information specialists, Saracevic and Covi specify the two communities as *Research Community*, e.g. academics, computer scientists and *Practice Community* e.g. information scientists, library professionals. These two groups are also recognised by Fuhr et al. (2001) and referred to as *Research Community*, and *Traditional Library Community*. It is argued in both papers Saracevic and Covi (2000) and Fuhr et al. (2001) that these 2 groups have different ideas of what makes a good digital library and also conflicting opinions on the Why? When? What? and How? requirements elicitation of evaluation. Saracevic and Covi begin to outline five elements that contain requirements for evaluation, "Construct, Context, Criteria, Measures and Methodology", however they state that there is no agreement on what should be specified in each element, and emphasise that more research needs to be done to resolve this. A challenge Fuhr et al. recognised and state that they hope their research goes some way to achieving.

It is clear from this and other aspects that Fuhr et al. have developed their conceptual framework based on Saracevic and Covi's framework as there are similarities in the basic conceptualisation. Both split the digital library into three dimensions, System, Collection and Users. This is in contrast to Blandford et al. who only conceptualised System and Users for CASSM and did not model Content as a separate dimension. The two frameworks vary in how they conceptualise the attributes within these domains. Saracevic and Covi propose a concept of levels of evaluation, where attribute groups are created that are split into two categories, User Centred, which contains Level 1, Social, Level 2, Institutional and Level 3, Individual and System Centred which contains, Level 5, Engineering, Level 6, Processing and Level 7, Content. These categories are sandwiched with Level 4, Interface. Each attribute group will have attributes assigned to it that will be used in evaluation to highlight usability problems. Some of these attributes may be common to all digital libraries and some will be unique, making it difficult to conceptualise a definitive list. It is my opinion that this system has been conceptualised in a similar way to how systems are modelled using UML based upon elicited requirements. They have modelled the system boundary, where the system ends and the user interaction begins. This may allow for integration of this framework into the design stage of the system development lifecycle, which could provide a valuable utility. In contrast Fuhr et al. do not use the concept of levels, they define attribute groups inside the three dimensions. Users contains the attribute groups Who, What, How and Why which considers the user interaction with the system from a user perspective. System contains the attribute groups User, information Access, System Structure and Document, which considers user interaction from a system perspective. Collection contains the attribute groups Content, Meta-Content and

Management, this dimension can potentially model how information scientists conceptualise the system. Each of these attribute groups contain a further set of attributes which will be assigned values in an evaluation, the author recommend that these values will either be binary yes or no responses or a restricted set of values. A further difference to Saracevic and Covi's framework is that Fuhr et al. attempt to provide a definitive list of attributes that can be evaluated through conceptualising possible usability problems.

The main difference between Fuhr et al. and Saracevic and Covi's research is the goal of their research, Fuhr is more concerned with information retrieval aspect of digital libraries as opposed to the library as a whole and this is clear in their arguments. This specialism in Information Retrieval provides a research requirement of creating test beds for new evaluation techniques to be tested on. They highlight the lack of benchmark data for digital libraries as a problem when comparing different frameworks. They hope their framework will provide a solution to the two research areas, by creating a classification scheme so all libraries can be compared attribute by attribute to form a test suite of libraries, these attributes can be evaluated individually and comparisons made. Their conceptual framework can be developed to form solutions to both classification and evaluation.

3. Section C - Tsakonas Framework vs. Sandusky Framework

Does the Tsakonas Framework highlight more usability problems than The Sandusky Framework, when evaluating the usability of multiple digital libraries?

The focus of this section of the review is to compare the Tsakonas Framework and the Sandusky Framework both with each other, but also to draw comparisons to the conceptual frameworks proposed by Fuhr et al. (2001) and Saracevic and Covi (2000) to establish which concepts have been developed to create these frameworks. Sandusky references both papers, where as Tsakonas et al. reference only Fuhr et al. Both frameworks have developed other techniques from different researchers that fall outside the scope of this review, which will contribute to distinguishing them.

The Tsakonas Framework models user interaction, and is concerned with how to evaluate user interaction with digital libraries. They discuss the problem highlighted by Blandford et al (2004), Saracevic and Covi (2000) and Fuhr et al (2001). in that there are different types of user and depending upon their specialism (or lack of) their interaction with the system will be different. They qualify the need for their research in relation to the difficulty in modelling users' perception of an interaction with relation to their information need, they also discuss how terminology needs to be evaluated to ensure it can be understood by all users and not be too specific to one user group. In contrast the Sandusky framework is concerned with creating techniques to evaluate the usability of multiple digital libraries in order to draw comparisons, but also to use these techniques on a single library to analyse cause and effect between attributes, how a usability problem in one attribute can cause a problem with another attribute and whether fixing one usability problem will fix another or cause more problems. Sandusky also describes the goal of establishing a standard vocabulary that can be used to categorise different digital libraries. He hopes his framework will achieve this by identifying general attributes that digital libraries have in common which will enable analysts to evaluate any digital library with the attributes providing points for comparison. The standard vocabulary will be drawn from common attributes of multiple digital libraries.

Both frameworks are similar in that they have developed Saracevic and Covi and Fuhr et al's conceptualisation of the Digital library in dimensions and attributes, however whereas the Tsakanas framework adheres closely to Fuhr et al's concepts of three dimensions, Sandusky proposes a six dimensional framework. Both frameworks share the concepts proposed by both Saracevic and Covi and Fuhr et al. of creating attribute groups within these dimensions and assigning attributes to be evaluated to an attribute group. Both frameworks propose a set of attributes but state that only with continued testing of the framework will a definitive list be able to be created.

The Tsakonas Framework specifies three concepts of interaction that will be evaluated, Performance, Usability and Usefulness of the system. These interactions occur within the three dimensional system modelled in line with Fuhr et al's concepts of System, User and Content. In addition, the Tsakanas framework splits the system into three sub systems, Interface, Information Retrieval and Advanced Functionality. This holds comparison to Saracevic and Covi in modelling the boundaries of the system and appreciation that in terms of interaction, the interface is the gateway to the functionality of the system. The authors state a difficulty in modelling Advanced Functionality as computer scientists and information scientists have different opinions as to where the boundary is between basic and advanced functionality. Similarly to Fuhr et al. evaluation requirements will be elicited in order

to ascertain the attributes for each interaction, these will be defined as criteria, it is these criteria that form the attributes to be evaluated. Methods are also modelled, these are the usability evaluation techniques that are best suited to highlighting usability problems for a particular set of attributes. Similarly to Blandford et al. (2004) CASSM, relationships can be defined between the interaction concepts, the system dimensions and criteria, e.g. Content (dimension) is related to Usefulness (interaction) and User (dimension) related to Information Need (criteria).

The Sandusky framework consists of six dimensions, Audience (users), Institution (e.g. university, company), Access (e.g. who, why, what, how), Content (documents), Services (advanced functionality) and Design and Development (e.g. building and maintaining, software lifecycle). Attributes are assigned to these dimensions, similarly to the Tsakanas framework, evaluation requirements will be elicited to define the attributes. Sandusky argues against the concept of levels proposed by Saracevic and Covi, stating that by tying attributes to be either user or system centred will inhibit the ability to compare how a usability problem in one attribute is affected by another. Therefore attributes relating to an evaluation requirement of a user interaction or a system response can be modelled in the same dimension, this is starkly different to the Tsakanas framework and to the concepts proposed by Saracevic and Covi and Fuhr et al. where there was a clear distinction between System, User and Content.

Tsakonas et al. highlight further developments of their framework will include more focus on performance. They state that they are not aiming to model system performance, but user performance with the system i.e. if a user can complete a task, in what time and whether or not they are satisfied with the results. This is very difficult to measure, in particular user satisfaction, further research will continue in eliciting attributes within the framework that will enable an evaluation to measure user's performance. A further problem they highlight is that that until there is sufficient benchmark data available for digital library research, they cannot produce a definitive set of attributes for their framework. The framework reported developed criterion based on data in the TREC test bed (www.trec.nist.gov) namely user effectiveness, user efficiency and user satisfaction, but state that as this data is relevant for information retrieval it is not specific enough for Digital library research.

Sandusky recognises the limitations of his framework if attributes are not evaluated in correlation with developers of the system, because it will be difficult to get access to the development methods. This is a common problem of digital library evaluation, discussed by Hartson et al. (2004), most usability evaluations are carried out after deployment of the system, rarely as part of the software development lifecycle and by usability specialists rather than designers, developers or users of the system. The author hopes to develop the framework further, incorporating more conceptualisations from referenced researchers in order to give the framework more weight in the academic world, whereby it may aid development of a standard set of tools for usability evaluation of digital libraries.

Unfortunately with both frameworks the research is theoretical, no experiments have been reported testing the effectiveness of their frameworks at highlighting usability problems when evaluating digital libraries. Whereas Sandusky reports that his framework is currently being tested commercially, Tsakanos et al. argue that until there are adequate test suites available they cannot accurately test the effectiveness of their framework at evaluating user interaction. Therefore, by review alone it is impossible to compare whether one framework highlights more usability problems than another, as there are no experiment results to compare.

4. Section D – Research Questions

4.1 General Research Question:

Which techniques are most appropriate for evaluating the usability of Digital Libraries?

Which usability evaluation technique is most appropriate in evaluating a user's interaction with a digital library system? Are the standard tools e.g. Heuristic Evaluation as proposed by Neilson substantial enough to highlight all usability problems in digital libraries? Blandford et al. (2004) analysed four usability evaluation techniques, but did not find any one technique to be better than another. There is not enough research in usability evaluation of digital libraries – user activity on digital libraries cannot be treated the same as a website or piece of software, although there are some similarities e.g. searching, browsing, information retrieval, meaningful text. Saracevic (2004), Hartson et al (2004).

A research goal has been established to find a standard technique to evaluate the usability of digital libraries as part of software development lifecycle, that will highlight an exhaustive list of usability problems and perhaps lead to standards for digital library design and development.

4.2 Specific Research Question:

Will evaluating digital libraries Attribute by Attribute allow for a comparison of usability problems of multiple digital libraries?

Will a technique of evaluating the usability of digital libraries attribute by attribute prove to be an effective standard tool for highlighting usability problems and move toward a "Toolbox of Techniques" Blandford (2004) for evaluating digital libraries? Can the technique be adapted into the software development lifecycle to ensure common usability problems are fixed in the design / implementation stage to reduce system development costs? Will evaluating digital libraries attribute by attribute allow multiple digital libraries to be evaluated and comparisons drawn, to determine a common set of usability problems generic to digital libraries that will allow for standards to be developed?

4.3 Highly Specific Research Question:

Does the Tsakonas Framework highlight more usability problems than Sandusky Framework, when evaluating the usability of multiple digital libraries?

The Tsakanos framework and Sandusky framework have developed concepts proposed in attribute by attribute evaluation research, extending the research to move closer to a technique for effectively evaluating the usability of multiple digital libraries. Will comparison of the two frameworks to see which highlights more usability problems be beneficial in achieving the goal of creating a "toolbox of techniques"? Will one framework stand out alone as the more effective tool for evaluating digital libraries and enable further research incorporating the framework into the software development lifecycle? Do these frameworks (or aspects of them) have the potential to be developed into a standard tool for evaluating multiple digital libraries? How do the two proposed frameworks compare at highlighting usability problems to the standard technique of Heuristic evaluation? This is important as

Heuristic Evaluation is seen as an industry standard for usability evaluations of any system.

A further research question that has been highlighted from this review is the potential to develop test suites of digital libraries to enable the testing and comparison of evaluation frameworks. This will be beneficial to research for creating a set of standard techniques for evaluating multiple digital libraries so that standards can be formed for accessibility and usability.

5. Section E – Comparison of Research Papers

Table showing comparison of papers used in section c

Authors	Research Question	Primary or Secondary	Research Methodology
TSAKONAS, G. KAPIDAKIS, S. PAPATHEODOROU, C.	1. A framework to evaluate user interaction with digital libraries. 2. Proposed method to compare digital libraries attribute by attribute to create test suites for testing digital library research	Primary (because they are proposing a new idea – even though no experiments)	Theoretical
SANDUSKY, R. J.	1. A framework to evaluate the usability of multiple digital libraries. 2. Can comparing differences and similarities between usability problems in digital libraries provide a standard vocabulary for future usability evaluations of digital libraries?	Primary (because they are proposing a new idea – even though no experiments)	Theoretical

6. Section F – Slideshow

6.1 Slideshow Describing Research Review Findings.

A slideshow has been created which summarises the research and findings. A printed slideshow of these slides can be found in Appendix A.

7. Section G - MSC Project Proposal

7.1 MSC Project Proposal Form

Name Stacey Greenaway
Student Number 0487622
Award Msc Computer Science (Internet Technology)
Address 9 School Street, Sedgley, Dudley, DY3 1HU
e-Mail (home, work, university) stacey.greenaway@blueyonder.co.uk
Telephone no(s) 01384 356798

Project Title Attribute by Attribute Usability Evaluation for Digital libraries.

Aim
To see if one framework highlights more usability problems than the other and if both/either/neither/ are more effective than heuristic evaluation – a usability evaluation standard.

Deliverable
Usability evaluations of two digital libraries and an analysis of the results comparing the effectiveness of both frameworks to each other and to heuristic evaluation. I have chosen only two digital libraries to evaluate in order to keep the project manageable in the 12 week timeframe.

Research question(s)
Does the Tsakonas Framework highlight more usability problems than Sandusky Framework, when evaluating the usability of multiple digital libraries?
(See above sections 4.3 for more detail)

Background
Began to research into usability evaluation methods for digital libraries, this highlighted a research area of evaluating attribute by attribute, this is an ongoing research area and as yet there is no technique accepted as the standard for performing usability evaluations attribute by attribute for digital libraries.

Method
Research further into whether test suites/benchmark data exists for digital library evaluations, if so method would change to use the test suites if possible.

This method is based upon no such test suites existing, which is my current understanding.

- Decide upon two digital libraries to evaluate.
- Gain an understanding of the functionality of the digital libraries
- Perform some test evaluations of other digital libraries to gain a familiarity with the evaluation techniques.
- Using the concepts of each framework elicit attributes for each digital library.
- Perform the evaluations of the 2 digital libraries:
  - Firstly – evaluate using heuristic evaluation to provide a control to compare the results to.
  - Secondly - evaluate using the Tsakanos framework.
  - Thirdly – evaluate using the Sandusky framework.
- Evaluate the results.

Evaluation

By using Heuristic evaluation method to evaluate the example digital libraries it will perform a control on which to base what is a good amount of usability problems to highlight. This is because this is proved to be a good standard for finding usability problems in the majority of systems. Theoretically, the two conceptual frameworks should out perform heuristic evaluation if the research in the literature review is correct. The success of Tsakanos and Sandusky frameworks at highlighting usability problems can be compared to firstly whether or not they highlight more than Heuristic evaluation and secondly if they highlight more or less than each other. It will probably be necessary to gather statistical analysis of the results to determine conclusively which technique is more effective.

Provisional schedule

Wk 1-2 – research different digital libraries and select four, two for learning evaluation skills and two for final evaluations. Familiarise myself with their functionality
Wk 2-4 – Perform test evaluations
Week 5-8 – Elicit attributes and evaluation tasks, write up process for final report. Make notes for evaluating the frameworks adaptability in this process.
Week 7-9 – perform the evaluations.
Week 10 – 11 Evaluate the results continue writing up report.
Week 12 – finish writing up report and evaluate the project.

Initial reading list
Please see references and bibliography (sections 9 and 10 of this report.)

Hardware/software needed
PC/laptop, access to internet, web browser.

Supervisor (if known)
Prof. Mike Thelwall

Client
Prof. Mike Thelwall

External contact (external to the school)

If there is, an external agreement form must be filled in (appendix B)

n/a

Will the project require interviewing anybody other than members of SCIT staff?	**Yes/No**
Will the project be using a questionnaire to gather data or evaluate the software/system?	**Yes/No**
Will the project involve human test subjects?	**Yes/No**
Will the project involve access to personal information on individuals?	**Yes/No**

Note, any project involving people requires ethical approval, that is, if you have answered yes to any of the above four questions. Thus any such project (in which interviews or questionnaires are to be used) form SC035 and SC036 MUST be completed (see Appendix E - Ethical Approval of Project Proposals).

8. Section H - Evaluation

8.1 Suitability To Conduct Research.

As I am inexperienced at usability evaluations, I may not be able to achieve an accuracy in results as that of an experienced evaluator. To account for this I will need to do some trial evaluations with the three methods to get more experience and familiarity with the processes, to highlight any areas that I need to do more research on, or that may need more time allowed. I also need to become fully familiar with the digital libraries I will be evaluating and what functionality they offer. I can then draw up tasks that will be evaluated and ensure all sections are covered. I may not realistically be able to access all areas of the digital library, as I will not have permission or full understanding of the system without contact to the developers. It may be an option that I can produce a more thorough evaluation by contacting the digital library administrators and/or developers to gain more information on existing usability evaluations or site map of the system. However, as a student project, it may be the case that they will not reply to me, perhaps they are not working on the digital library project anymore and therefore do not have time to reply or can not remember certain aspects of the project.

8.2 Process Evaluation.

The process I have gone through to create this literature review has been highly beneficial in equipping me with skills to carry forward into my MSc Project. I have learned valuable lessons that will improve my efficiency and time management when I undertake the MSc project. Without the homework assignments, I would have had more problems with time management, they gave a weekly direction to the assignment, which can be adapted into other research projects. One main problem I had was keeping focussed on the assignment and not going off in tangents following interesting citation chains or links in search results, that were not strictly relevant to digital libraries. I learned early on in the project to keep focussed on digital libraries and not to stray into the extremely broad topic of usability evaluation, this enabled me to narrow my research area considerably. It did however create another problem in that there was less research on usability evaluation techniques for digital libraries than I expected. It took deeper reading to find out about the attribute by attribute method. I found it hard to reject papers based on abstract only, I had to read a little deeper sometimes to find out they were not completely relevant to the topic. It was frustrating when you would find a paper, which based on the abstract was perfect, but then it would be a poster session or tutorial, or a journal article I could not get access too. In future projects I need to begin critically reading and making notes about the papers earlier rather than relying on skimming, which invariably leads to missing out certain aspects of the research and maybe wrongly rejecting or accepting the research paper as valid to the topic. A valuable lesson I have re-learned and will not ignore again is to make a note of the full Harvard Reference for a paper as it is found, even if the paper is then rejected it will save a lot of time in the long run. One aspect of the process that could be deemed both good and bad is the quality of the research I found, as it gave me a lot to write about, I had to overwrite the recommended word counts in order for the review to flow and for the specific research areas to be appropriately analysed. The research process as a whole went well, early on in the process I found a paper, Blandford et al. (2004) which inspired my research questions and focussed my research toward attribute by attribute evaluation techniques. Although there was not a wealth of research available, I believe I found research papers that helped to form an interesting and valid proposal for a research project.

9. References

ADAMS, A. & BLANDFORD, A. (2005) Digital libraries' support for the user's Information Journey. In Proc. ACM/IEEE joint conference on Digital libraries 2005. pp.160-169.

BLANDFORD, A. (2004) Understanding Users' Experiences: Evaluation of Digital Libraries. *in Proceedings of the Delos workshop on evaluation of Digital libraries*, *Padova, October 2004*, pp. 31-34.

BLANDFORD, A. CONNELL, I. EDWARDS, H. (2004b) Analytical Usability Evaluation For Digital Libraries: A Case Study. *in Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries Tuscon, AZ.* New York, NY, USA: ACM Press, pp. 27–36

BLANDFORD, A. KEITH, S. FIELDS, B. (2003) Tailoring Claims Analysis to the design and deployment of digital libraries: a case study. Available from:www.cs.mdx.ac.uk/ridl/UET/TR5Usingclaims.pdf

BLANDFORD, A. BUCHANAN, G. (2002) Workshop Report: Usability Of Digital Libraries @ Joint Conference On Digital Libraries '02. *in ACM SIGIR Forum,* **36**(2), pp. 83 – 89.

BLANDFORD, A. STELMASZEWSKA, H. BRYAN-KINNS, N. (2001) Use of multiple digital libraries: a case study. *in Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries Roanoke, Virginia, United States*, New York, NY, USA: ACM Press, pp. 179 – 188.

CAIDI, N. KOMLODI, A. (2003) Digital libraries across cultures: design and usability issues outcomes of the "cross-cultural usability for digital libraries" workshop at Joint Conference On Digital Libraries '03. in ACM SIGIR Forum, **37**(2), pp. 62-64.

FOX, E. A. AKSCYN, R. M. FURUTA, R. K. LEGGETT, J. J. (1995) Digital libraries. *In Communications of the ACM ,* New York, NY, USA: ACM Press, **38(**4), pp. 22-28 .

FUHR, N. HANSEN, P. MABE, M. MICSIK, A. SØLVBERG, I. (2001) Digital Libraries: A Generic Classification and Evaluation Scheme. *in Proceedings of the 5th European Conference on Research and Advanced Technology for Digital Libraries,* London, UK**:** Springer-Verlag, **2163** pp. 187-199.

HARTSON, H.R. SHIVAKUMAR, P. PÉREZ-QUIÑONES, M. A. (2004) Usability Inspection of Digital Libraries: A Case Study. *in International Journal on Digital Libraries*, Springer Berlin / Heidelberg, 4(2), pp. 108-123.

KEITH, S. BLANDFORD, A. FIELDS, B. THENG, Y. T. (2002) An investigation into the application of Claims Analysis to evaluate usability of a digital library interface. *in Proceedings of workshop on Usability of Digital Libraries at Joint Conference On Digital Libraries '02*. Available from www.uclic.ucl.ac.uk/annb/DLUsability/JCDL02.html

KIM, K. (2002) A Model-based Approach to Usability Evaluation for Digital Libraries*. in proceedings of Joint conference on digital libraries '02 Workshop on Usability of Digital Libraries*, p. 33.

KLAS, C. FUHR, N. KRIEWEL, S. ALBRECHTSEN, H. TSAKONAS, G. KAPIDAKIS, S. PAPATHEODOROU, C. HANSEN, P. KOVACS, L. MICSIK, A. JACOB, E. (2006) An experimental framework for comparative digital library evaluation: the logging scheme. *in proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries JCDL '06 Chapel Hill, NC, USA,* New York, NY, USA: ACM Press, pp. 308-309.

KYRILLIDOU, M. GIERSCH, S. (2005) Developing the DigiQUAL protocol for digital library evaluation. *in Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries, SESSION: Users and interaction track: understanding user needs and perceptions, Denver, CO, USA*, pp. 172–173.

MARSDEN, G. CHERRY, R. HAEFELE, A. (2002) Small screen access to digital libraries. *in CHI '02 extended abstracts on Human factors in computing systems, Minneapolis, Minnesota, USA,* New York, NY, USA: ACM Press, pp. 786-787

MCCRAY, A. T. GALLAGHER, M. E. (2001) Principles For Digital Library Development. *in Communications of the ACM*, 44(5), pp. 48–54.

REEVES, T. C. BUHR, S. BARKER, L. (2005) Evaluating digital libraries. *in proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries, Denver, CO, USA,* New York, NY, USA: ACM Press, pp. 420 – 420.

SARACEVIC, T. (2004) Evaluation of digital libraries: An overview. *in Proceedings of the Delos workshop on evaluation of Digital libraries*, *Padova, October 2004.*

SARACEVIC, T. COVI, L. (2000) Challenges for digital library evaluation. *In D. H. Kraft (Ed.), Knowledge Innovations: Celebrating Our Heritage, Designing Our Future. Proceedings of the 63rd Annual Meeting, November 11-16, 2000, Chicago, IL* Washington, D.C.: American Society for Information Science. pp. 341-350.

SANDUSKY, R. J. (2002) Digital Library Attributes: Framing usability research*. in Proceedings of the Workshop on Usability of Digital Libraries at Joint Conference On Digital Libraries,* p. 35.

THENG, Y. L. (1999) "Lostness" and digital libraries. *in proceedings of the fourth ACM conference on Digital libraries,* Berkeley, California, United States, New York, NY, USA: ACM Press, pp. 250 – 251.

THENG, Y. L. MOHD-NASIR, N. THIMBLEBY, H. (2000) Purpose and Usability of Digital Libraries. *in Proceedings of the fifth ACM conference on Digital libraries*, New York, NY, USA: ACM Press, pp. 238 – 239.

TSAKONAS, G. KAPIDAKIS, S. PAPATHEODOROU, C. (2004) Evaluation of User Interaction in Digital Libraries. *in proceedings of the Delos workshop on evaluation of Digital libraries 2004.*

10. Bibliography

ARMS, W. Y. AYA, S. DMITRIEV, P. KOT, B. J. MITCHELL, R. WALLE, L. (2006) Building a research library for the history of the web. *in Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries '06,* New York, NY, USA: ACM Press, pp. 95–102.

BRYAN-KINNS, N. BLANDFORD, A. (2000) A survey of user studies for digital libraries. RIDL Working paper, Available from www.cs.mdx.ac.uk/ridl/DL**user**.pdf

CHAMPENY, L. BORGMAN, C. L. LEAZER, G. H. GILLILAND-SWETLAND, A. J. MILLWOOD, K. A. D'AVOLIO, L. FINLEY, J. R. SMART, L. J. MAUTONE, P. D. MAYER, R. E. JOHNSON, R. A. Developing a digital learning environment: an evaluation of design and implementation processes. (2004) in Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries, Tuscon, AZ, USA, New York, NY, USA: ACM Press, pp. 37 – 46.

HEDMAN, A. (1999) Creating digital libraries together—collaboration, multimodality, and plurality.   in ACM SIGCSE Bulletin , Proceedings of the 4th annual SIGCSE/SIGCUE ITiCSE conference on Innovation and technology in computer science education ITiCSE '99, Cracow, Poland, New York, NY, USA: ACM Press, 31(3), pp. 147 – 150.

HERTZUM1, M. JACOBSEN2, N. E. (2001) The Evaluator Effect: A Chilling Fact about Usability Evaluation Methods. *in International Journal of Human-Computer Interaction*, 13(4), pp. 421-443.

HUAJING, L.   COUNCILL, I. G. BOLELLI, L. ZHOU, D. SONG, Y. LEE, W. C. SIVASUBRAMANIAM, A. GILES, C. L. (2006) Citeseerχ: A Scalable Autonomous Scientific Digital Library. in Proceedings of the 1st international conference on Scalable information systems InfoScale '06, Hong Kong, New York, NY, USA: ACM Press, 152(18).

JENG, J. H. (2004) Usability of digital libraries: an evaluation model. *in Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries, Tuscon, AZ, USA,* New York, NY, USA: ACM Press, p. 407.

JONES, S. PAYNTER, G. (1999) Topic-based browsing within a digital library using keyphrases.   in Proceedings of the fourth ACM conference on Digital libraries Berkeley, California, United States, New York, NY, USA: ACM Press, pp. 114 – 121.

LAWRENCE, S. GILES,C. L. BOLLACKER, K. (1999) Digital Libraries and Autonomous Citation Indexing. Los Alamitos, CA, USA: IEEE Computer Society Press, 32(6), pp. 67 – 71.

MAGNUSSEN, A. (2003) Creating Digital Libraries: A Model For Digital Library Development. in 10th Asia Pacific Special Health and Law Librarians Conference – Adelaide Aug 2003 pp. 24–27.

MEAD, J. P. GAY, G. (1995) Concept mapping: an innovative approach to digital library design and evaluation. ACM SIGOIS Bulletin, New York, NY, USA: ACM Press, 16 (2), pp. 10 - 14 .
TOMS, E. G. DUFOUR, C.   HESEMEIER, S.   (2004) Measuring the user's experience with digital libraries. in Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries Tuscon, AZ, USA, New York, NY, USA: ACM Press, pp. 51 - 52

SILVA, L. V. A. LAENDER, H. F. GONÇALVES, M. A. (2005) A Usability Evaluation Study of a Digital Library Self-Archiving Service. in Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries, SESSION: Users and interaction track: understanding user needs and perceptions, Denver, CO, USA, New York, NY, USA: ACM Press, pp. 176-177.

THONG, J.Y.L. HONG, W. TAM, K. Y. (2004) What leads to acceptance of digital libraries?, in Communications Of The Acm, 47(11), pp. 78 – 83.

WITTEN, I. H. BODDIE, S. J. BAINBRIDGE, D. MCNAB, R. J. (2000) Greenstone: a comprehensive open-source digital library software system. in Proceedings of the fifth ACM conference on Digital libraries, San Antonio, Texas, United States, York, NY, USA: ACM Press, pp. 113 – 121.

Appendix A - Slideshow

Usability Evaluation of Digital Libraries
Stacey Greenaway

Digital Libraries – General Research Areas
- Web based
- Metadata
- Automate
- Obtaining information
- Preservation
- Quality of Service
- Intellectual property rights
- HCI, Usability, Accessibility

Digital Libraries – HCI, Usability, Accessibility
- Accessible regardless of disability, language or cultural differences.
- Keyword searching.
- Ability to browse topics.
- Intuitive interface.
- Content optimised.
- Quick information retrieval.
- Good indexing (Metadata)

Usability Evaluation - Usability Problems
- No standards for digital library usability evaluation.
- Need to highlight common usability problems.
- How to model user behaviour?
- Users information need.
- Computer Scientist vs. Information Scientist
- Comparing multiple digital libraries

Usability Evaluation – Standard Techniques
General Research Question:
Which techniques are most appropriate for evaluating the usability of Digital Libraries?
- Usability Inspection.
- Heuristic Evaluation.
- Cognitive Walkthrough
- Claims Analysis.

None of the above techniques were found by Blandford et al. (2004) or Hartson et al. (2004) to be hugely successful at highlighting problems.

Usability Evaluation – Attribute by Attribute
Specific Research Question:

Will evaluating digital libraries Attribute by Attribute allow for a comparison of usability problems of multiple digital libraries?

3 proposed conceptual frameworks for evaluating digital libraries attribute by attribute.
- CASSM - (Concept-based Analysis of Surface and Structural Misfits) Blandford et al. (2004)
- Saracevic and Covi Framework (2000).
- Fuhr Framework (2001)

Usability Evaluation - Attribute by Attribute

Basic Method:

- System split into dimensions
  CASSM – 2 Dimensions – System & User
  Saracevic and Covi, Fuhr – 3 Dimensions – System, User & Content
- Attributes elicited and assigned to attribute groups within the dimension.
- Each attribute is evaluated
- Attributes can be analysed and compared to results of other evaluations.

Usability Evaluation - Attribute by Attribute

Does the Tsakonas Framework highlight more usability problems than Sandusky Framework, when evaluating the usability of multiple digital libraries?

Methods (additional to basic method proposed by Saracevic and Covi, Fuhr :

Tsakanos Framework:
- 3 Dimensions: System, User & Content.
- 3 sub systems: Interface, Information Retrieval and Advanced Functionality.
- Evaluate 3 conditions: Performance, Usability and Usefulness of the system.
- Define criteria (attributes) and methods (tools to evaluate the attributes)
- Relationships defined between the interaction concepts, the system dimensions and criteria.

Usability Evaluation - Attribute by Attribute

Does the Tsakonas Framework highlight more usability problems than Sandusky Framework, when evaluating the usability of multiple digital libraries?

Methods (additional to basic method proposed by Saracevic and Covi, Fuhr :

Sandusky Framework:
- 6 Dimensions – Audience, Institution, Access, Content, services and Design and Development. (referred to as attribute group)
- No separation or distinction between system, user and content tasks, individual attributes can be compared to each other easier.
- Evaluate cause an effect between attributes.

Usability Evaluation - Attribute by Attribute

Does the Tsakonas Framework highlight more usability problems than Sandusky Framework, when evaluating the usability of multiple digital libraries?

Comparison:

- Both frame works are conceptual.
- Both frame works do not provide experiments to help compare the efficiency of the frame works.
- Both state further development required.
- The 2 frame works share goal of evaluating the usability of multiple digital libraries, but in different contexts, for different purposes and have differing sub sets of research goals.
- Inconclusive results, experiments need to be conducted analysing effectiveness of both frame works on highlighting usability problems of a selection of digital libraries.

Usability Evaluation - Attribute by Attribute

Benefits:

- Comparing multiple digital libraries
- Find common usability problems – and fix them
- Produce standards for evaluation
- Create test suites for digital library research
- Compare cause and effect between attributes in a single digital library
- Tool for classification

Usability Evaluation - Attribute by Attribute

Negatives:

- Difficult to predict user behavior.
- Hard to model user satisfaction.
- No benchmarks for digital library research.
- Difficult to produce definitive set of attributes.
- No comprehensive checklist as all libraries different.
- Some attributes elicited at each evaluation.
- Requires system developers knowledge.

Usability Evaluation - Attribute by Attribute

Conclusion:

▶ Evaluating digital libraries attribute by attribute is a promising alternative to standard usability evaluation tools.

▶ More research needs to be done to test the strengths of these frameworks.

▶ More research creating test suites/benchmark data

▶ Requirement for standards in digital library development and evaluation.

▶ Through literature review alone it is impossible to determine whether any conceptual framework is more efficient than another.

Usability Evaluation - Attribute by Attribute

References:

BLANFORD, A. CONNELL, I. EDWARDS, H. (2004b) Analytical Usability Evaluation For Digital Libraries: A Case Study. in Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries Tuscon, AZ. New York, NY, USA: ACM Press, pp. 27–36

FUHR, N. HANSEN, P. MABE, M. MICSIK, A. SØLVBERG, I. (2001) Digital Libraries: A Generic Classification and Evaluation Scheme. in Proceedings of the 9 h European Conference on Research and Advanced Technology for Digital Libraries, London, UK: Springer-Verlag, 2163 pp. 187-199.

SANDUSKY, R. J. (2002) Digital Library Attributes: Framing usability research. in Proceedings of the Workshop on Usability of Digital Libraries at Joint Conference On Digital Libraries, p. 35.

SARACEVIC, T. COVI, L. (2000) Challenges for digital library evaluation. In D. H. Kraft (Ed.), Knowledge Innovations: Celebrating Our Heritage, Designing Our Future. Proceedings of the 63rd Annual Meeting, November 11-16, 2000, Chicago, IL. Washington, D.C.: American Society for Information Science, pp. 341-350.

TSAKONAS, G. KAPIDAKIS, S. PAPATHEODOROU, C. (2004) Evaluation of User Interaction in Digital Libraries. in proceedings of the Delos workshop on evaluation of Digital libraries 2004.