Evaluating Existing Digital Libraries as a Prototype with the Suggested Criteria: Content, Usability, and Performance Evaluation Criteria

(“Boaz”) Sunyoung Jin
Graduate School of Library and information Science
University of Illinois at Urbana-Champaign, Illinois, USA
sunjin@illinois.edu
Abstract

The proposed International Open Public Digital Library (IOPDL)\(^1\) is *open public digital library* for the public to access collections all over the world either free of charge or with fewest possible limitations as a non-profit organization (Jin). It will consist of new published collections and links to existing qualitative and quantitative collections of Well-Designed Digital Libraries (WDDLS) in many subject areas. To determine WDDLs, evaluation tools and methods were investigated. One combined method is suggested to evaluate multiple performances together of existing digital libraries with the suggested evaluation criteria: Content, Usability and Performance Evaluation (CUPE) criteria. The CUPE criteria include mainly three evaluations with seven sub-criteria: Content quality evaluation (with content quality criteria), Usability evaluation (with Accessibility, Convenience, Consistency and Visual design criteria), and Performance evaluation (with Response time and Relevance criteria).

In the Content quality evaluation, experts recommended sixty three digital libraries as candidates of WDDLs in fifteen subject domains (based on Library of Congress classification). Usability of the candidate is evaluated by using open sources for Accessibility evaluation, and by heuristic method for Interfaces’ usability evaluation. In the performance evaluation, the developed simulation programs evaluate again their response time and relevance (by word density). All seven evaluations with CUPE criteria are done with carefully considered check lists. The evaluation is limited regionally almost in the USA and timely on 2010. As a result, thirty four out of sixty three candidate digital libraries turn out as well-designed digital libraries in fifteen subject areas. Many problems of existing digital libraries are found, which will encourage their developments. The suggested CUPE is one of the appropriate ways to determine well-designed digital libraries, to evaluate existing digital libraries in many subject areas, and to point out their problems in content quality, usability and performance. The evaluations are effective and efficient to assess overall their various performances together and many digital libraries at once. The evaluation methods may further suggest a certain way to evaluate national digital libraries all over the world for the near future.

\(^1\) National Open Public Digital Library (NOPDL) was proposed in 2008. The NOPDL is renamed into International Open Public Digital Library (IOPDL) in 2013.
# Table of Contents

Abstract ......................................................................................................................................................... a

I. Introduction ........................................................................................................................................... 1

II. Existing Evaluation Criteria and Methods ............................................................................................ 1
  1. Usability Criteria .................................................................................................................................. 2
  2. The Performance Criteria ................................................................................................................ 3
  3. Problems of Existing Evaluation Criteria (Why do we need New Criteria?) ................................... 4

III. The Suggested Evaluation Criteria: ‘Content, Usability and Performance Evaluation’ (CUPE) Criteria ........................................................................................................................................ 4
  1. The CUPE Criteria ............................................................................................................................ 6
  2. The Content Evaluation Criteria ....................................................................................................... 6
  3. Problems of Existing Evaluation Criteria (Why do we need New Criteria?) ................................... 4
  4. The Performance Evaluation Criteria ............................................................................................... 8

IV. The Content Evaluation .................................................................................................................... 9
  1. Methodology ................................................................................................................................... 10
  2. The Content Evaluation Result ....................................................................................................... 11
  3. The Analyses and Discussion of the Content Evaluation ............................................................... 13

V. The Usability Evaluation .................................................................................................................... 14
  1. Accessibility Evaluation ...................................................................................................................... 14
     1-1. Methodology ............................................................................................................................. 14
     1-2. Results and Analyses ............................................................................................................... 16
     1-3. Total Analyses with Seven Accessibility Tools ...................................................................... 19
  2. Interface Usability Evaluation .......................................................................................................... 22
     2-1. Methodology ............................................................................................................................. 22
     2-2. Results and Analyses of Convenience/ Ease of Use Evaluation ............................................. 23
     2-3. Results and Analyses of Interface Consistency Evaluation ..................................................... 26
     2-4. Results and Analyses of Visible Design and Aesthetic Appeal Evaluation ............................. 27
  3. Analyses of Interface Usability Evaluations with Three Criteria .................................................... 28

VI. The Performance Evaluation ............................................................................................................. 31
  1. The Response / Retrieval Time Evaluation ..................................................................................... 31
     1-1. Methodology ............................................................................................................................. 31
     1-2. The Result of the Response Time Evaluation .......................................................................... 33
     1-3. The Analyses of the Response Time Evaluation ................................................................... 34
  2. The Relevance Evaluation ................................................................................................................. 35
     2-1. Methodology ............................................................................................................................. 35
     2-2. The Results of Relevance Evaluation ....................................................................................... 37
     2-3. The Analyses of the Relevance Evaluation ............................................................................. 37
  3. The Ranking by the Performance Evaluations .................................................................................. 38

VII. Total Results of Seven Evaluations ................................................................................................ 39
  1. Methodology to Combine Seven Evaluations ................................................................................. 40
  2. Total Combined Results of Seven Evaluations .............................................................................. 41
  3. Total Analyses of Seven Evaluations .............................................................................................. 42
  4. Well-Designed Digital Libraries ..................................................................................................... 43

VIII. Conclusion and Discussion ............................................................................................................. 46

Reference ................................................................................................................................................... 47

Appendix A - Results of All Seven Evaluations with the CUPE criteria ................................................... 49
Table of Figures

Figure 1: The Result of United States Department of Defense by FAE .........................................................16
Figure 2: The Result of United States Department of Defense by Cynthia Says ..............................................17
Figure 3: Averages of Evaluation Tools Figure 4: Averages of Accessibility in Subject Domains ..................20
Figure 5: Showing clustered related information Figure 6: Showing Help Documents .............................24
Figure 7: The browse web page of SMC. Figure 8: Showing clustered related information .......................25
Figure 9: Showing clustered related information Figure 10: Showing Help Documents ............................25
Figure 11: Showing Consistency between data entry and data display ......................................................26
Figure 12: Showing inconsistency between data entry and data display ....................................................27
Figure 13: Comparing Averages of Interfaces Usability Evaluations in Fifteen Subject Areas .....................29
Figure 14: Comparisons of Total Averages among Fifteen Subject Areas ...............................................29
Figure 15: The Response Time Rates Figure 16: Averages of Response Time ............................................35
Figure 17: The Relevance Rates, based on the scores Figure 18: Averages of Relevance Rates, .................38
Figure 19: The Percentages of the final average of all candidate digital libraries .....................................42
Figure 20: Comparing the Average based on Subject Domains ..............................................................43

Table of Tables

Table 1: The selected evaluation tools ..............................................................................................................15
Table 2: The Highest Score Digital Libraries by Accessibility evaluation in each Subject Domain ...............22
Table 3: The Ranking by Interface Usability Evaluations with three Usability Evaluation Criteria ....... .31
Table 4: The Ranking by the Performance Evaluations ..............................................................................39
Table 5: International Children’s Digital Library scores ...........................................................................41
Table 6: Sample Results of All Evaluations with the CUPE criteria ...........................................................42
Table 7: Well-Designed Digital Libraries based on Scores ...........................................................................46

Table of Equation

Equation 1: Average Response Time ...........................................................................................................31
Equation 2: Average Link Response Time ....................................................................................................32
Equation 3: Average Search Response Time ................................................................................................32
Equation 4: An Average Response Time for Response time evaluation ....................................................32
Equation 5: Keyword Density .....................................................................................................................36
Equation 6: Relevance Rate ........................................................................................................................36
Equation 7: Final Average for Each Digital Library ....................................................................................40
Equation 8: Total Average of All Candidate Digital Libraries ..................................................................41
Equation 9: Final Averages of Well-Designed Digital Libraries .................................................................41
I. Introduction

These days, people tend to access information or books online via digital libraries. However, information access through digital libraries is limited by factors relating to geography, time, cost, authorization, and quality and quantity of collections (Jin). That is, to date, there is no worldwide international public digital library which presents a solution of above pointed limitations. Solutions to the problem, International Open Public Digital Library (IOPDL) is proposed (Jin). The purpose of the IOPDL is for the public to access information at free charge without limitations or with fewest possible limitations. One method of establishing the IOPDL is for existing well-designed digital libraries all over the world to cooperate and consist of it. This approach will reduce budget, time, and effort as well as encouraging WDDLs to improve their exclusive contents considerably (Jin). Well-designed libraries (WDDLs) are defined as qualified libraries by usability, performance, and content evaluations in their subject domains.

The paper demonstrates a prototype that conducts the general requirement, determining well-designed digital libraries (WDDLs), to establish the proposed IOPDL. For it, evaluations of existing digital libraries are essential in many subject areas. Through investigations of existing evaluation criteria and methods, the main finding is that there is no appropriate method to evaluate several performances together. Existing methods evaluate a specific field of many performances. Since WDDLs are defined by content quality, usability and performance evaluations, new criteria, “Content, Usability, and Performance Evaluation (CUPE) Criteria,” are suggested, which are illustrated in the chapter III. From chapter IV, actual evaluations of content, usability, and performance are described with detail explanations: which criteria and check lists are used for each evaluation, which evaluation methods are used, and results and analyses of the results. The chapter IV discusses the content evaluation. Through the content evaluation, sixty three digital libraries are recommended by experts as the candidates of WDDLs in fifteen subject areas (based on Library of Congress classification). The chapter V discusses the Usability evaluation of the candidate of WDDLs with Accessibility, Convenience/ ease of use, Interfaces’ consistency, and Visible design and Aesthetic Appeal evaluation criteria. The chapter VI discusses the Performance evaluation with simulation programs to measure response/ retrieval time and relevance of query results. The chapter VII discusses total evaluation results of all seven evaluations analyzing them. Finally, the turned out WDDLs are listed. The evaluation is limited regionally almost in the USA and timely on 2010. The further evaluations for national digital libraries all over the world remain. The prototype will guide a certain way to evaluate globally national digital libraries all over the world.

II. Existing Evaluation Criteria and Methods

Evaluation of digital Libraries is defined as “a fact finding, evidence based value measuring, integrated in the management process of digital libraries” (Tammaro & Marlino). Evaluating digital libraries is based on criteria. To find the defined well-designed digital libraries, we need particular criteria to evaluate multiple performances together of existing digital libraries. For it, existing evaluation criteria and methods are investigated whether we can use them appropriately for our evaluation purposes, or we need new criteria that fit the purpose and goal of the project.

Borgman (2003) explains at least four types of evaluation, according to stages of evaluations:

“Formative evaluation begins at the initial stages of development project to establish baselines on current operations, set goals, and determine desired outcomes.

- Summative evaluation takes place at the end of a project to determine if the intended goals were met. Goals and outcomes must be compared to initial states.

- Iterative evaluation takes place throughout a project, beginning in the earliest design and development stages. Interim stages of design are assessed in comparison to design goals and desired outcomes, and the results inform the next stages of design.
Comparative evaluation requires standardized measures that can be compared across systems. Communities can identify and validate measures” (Borgman, 2003, p. 6).

Since evaluating existing digital libraries is not a development of a project, comparative evaluation that compares across systems with standardized measures may be appropriate for the project. Saracevic (2000) states what we should consider in evaluations: What should we evaluate? For what purpose do we evaluate? Who should evaluate? At what level do we evaluate? Upon what criteria do we evaluate? How to apply them in evaluation? (Saracevic, 2000, p. 1). Saracevic summarizes a good number of evaluation criteria, standardized measurements, for libraries:

“Traditional library criteria
- Collection: purpose, subject, scope, authority, coverage, …, preservation, persistence;
- Information: accuracy, appropriateness, …, ownership;
- Use: accessibility, availability, searchability, usability

Traditional IR criteria
- Relevance (leads to measures of precision and recall),
- Satisfaction, success; and
- Index, search, output features

Traditional human-computer interaction/interfaces criteria
- Usability, functionality, effort,
- Connectivity, reliability;
- Design features;
- Navigation, browsing; and

In those criteria, particularly, Saracevic points out that Usability is being broadly used in digital library evaluation as traditional systems evaluation criteria. The performance evaluation examines effectiveness, “How well does a system perform?” and efficiency “At what cost” (Saracevic, 2004). What are usability and performance evaluations? They are explained in this chapter.

1. Usability Criteria

What is Usability?
ISO defines usability as “the effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments” (ISO), where:
- Effectiveness means "the accuracy and completeness with which specified users can achieve specified goals in particular environments;"
- Efficiency means "the resources expended in relation to the accuracy and completeness of goals achieved";
- Satisfaction means "the comfort and acceptability of the work system to its users and other people affected by its use" (ISO).

What is Usability Evaluation?
In Usability.gov website, usability evaluation is generally referred to “how well users can learn and use a product (e.g. a website, a software application, or a digital library) to achieve their goals and how satisfied they are with that process” (Usability.gov). Hilbert and Redmiles define usability evaluation as “the act of measuring (or identifying potential issues affecting) usability attributes of a system or device with respect to particular users, performing particular tasks, in particular contexts” (Hilbert & Redmiles, 2000).

The usability testing goal is to identify any usability problems on participants’ performance such as time on task and error rates. The methods to evaluate usability are Heuristic Evaluation, Card Sorting, Contextual Interview, Focus Groups, Survey, Prototyping, and Use Cases (Usability.gov).
What does Usability measure?

There are common measurements for usability testing. According to Usability.gov, common five factors are measured in usability testing:

- "Ease of learning (how fast can a first user learn it satisfactorily well),
- Efficiency of use (once an experienced user has learned to use the system, how fast can he or she accomplish tasks?),
- Memorability ("if a user has used the system before, can he or she remember enough to use it effectively the next time or does the user have to start over again learning everything?),
- Error frequency and severity (errors rate, and how do users recover from these errors?),
- Subjective satisfaction" (Usability.gov).

Fleming (1998) explains important properties for website usability related to navigability:

- "Consistency of presentation and controls across the site, natural organization of information (clear structure, systematic labels, clear and meaningful labels);"
- "Contextual navigation, in terms of environment, "type" of users, particular devices, etc.; someone often considers also how much information is given for providing a context to the user (where is he, where he can go, and so on);"
- "Robustness, (i.e. how well the Web site handles technology used by users that has not been foreseen by developers);"
- "Flexibility (e.g. availability of graphic and textual versions, redundant indexes and site maps, duplicated image map links)” (Fleming, 1998).

In detail, Saracevic (2000) lists usability criteria applied in various studies:

- For content of a portal or site: “accessibility, availability, clarity, complexity of organization or structure, adequacy, coverage, quality, accuracy, authority and reliability.
- For process carrying out tasks (e.g. search, navigation, browsing, finding, evaluation or obtaining a resource): learnability, convenience, ease of use, confusion, completion, interpretation difficulty, and error rate.
- For format: attractiveness, consistency, representation of labels and messages’ communicativeness.
- For overall assessment: satisfaction, relevance, usefulness of results, barriers, quality of experience, learning, and preferences” (Saracevic, 2004, p. 6).

2. The Performance Criteria

What are Performance Evaluation criteria?

Saracevic explains that the performance evaluations can assess “effectiveness (How well does a system (or any of its parts) perform that for which it was designed?), efficiency (At what cost (costs could be financial or involve time or effort)?), and a combination of these two (i.e., cost-effectiveness)” (Saracevic, 2000). Saracevic lists systems-centered evaluation criteria to assess digital libraries:

- “For technology performance: response time, processing time, speed, and capacity.
- For process/algorithm performance: relevance of obtained results, clustering, similarity, flexibility, error rate, retrieval time, path length and optimization.
- For overall system: maintainability, scalability, interoperability, sharability, and costs” (Saracevic, 2004, p. 7).

Fuhr et al. (2007) explain relevance of performance evaluation. Relevance focuses on the interaction “between users, usage, information needs and informational content” (Fuhr, Tsakonas, & Aalberg, 2007). Relevance of system performance is measured by recall and precision, which define model completeness and precision of systems (Fuhr, Tsakonas, & Aalberg, 2007).
3. Problems of Existing Evaluation Criteria (Why do we need New Criteria?)

As Saracevic points out, the main problem of existing evaluation criteria is that there is no uniform definition and a standard method. There are no standardized criteria for diverse purposes. Depending on different evaluation goals and audiences, different approaches and criteria have applied appropriately. Thus, while a criterion and method are appropriate to a digital library, they cannot be also useful for another digital library evaluation.

Since a digital library is a complex system, it cannot be evaluated by one or two criteria. Each criterion evaluates a part of complex system. For illustrate, although many usability and performance evaluation criteria have been developed, each criterion has limitations to evaluate one or two specific performance not multiple performances together. But, usability and performance cannot be evaluated by one or two criteria. We need more combined usability and performance evaluation criteria that can evaluate multiple performances of digital libraries together.

Moreover, there are no specific evaluation criteria that evaluate content quality of collections. For our purposes, we need to evaluate content quality of digital libraries, because well-designed digital library (WDDLs) should possess quality collections in a subject area as representative qualitative WDDLs.

Therefore, based on our evaluation goal to find WDDLs, we need new criteria to evaluate multiple performances of digital libraries. Since there are no appropriate criteria to conduct evaluations for the project, thus, I suggest a combined evaluation criteria that evaluate content, usability and performance of existing digital libraries.

III. The Suggested Evaluation Criteria: ‘Content, Usability and Performance Evaluation’ (CUPE) Criteria

Through the investigation of existing evaluation criteria, the main finding is that there is no standard way to evaluate multiple performances of many digital libraries. Our purpose to evaluate digital libraries is to find well-designed digital libraries that satisfy the presented requirements. Thus, we need particular criteria to evaluate multiple performances (e.g., content quality, usability and performance) of existing digital libraries. The criteria should evaluate quality of collections in their subject domain. The criteria should evaluate their usability in various viewpoints. Lastly, the criteria should evaluate their various performances. Each criterion should evaluate varied digital libraries in depth and width. Hence, the suggested criteria are named as ‘Content Usability & Performance Evaluation (CUPE)’ criteria, which is a set of criteria. Before defining the CUPE criteria for our project evaluations, first of all, we should consider “for what purpose do we evaluate? what should we evaluate? who should evaluate? at what level do we evaluate? upon what criteria do we evaluate? how to apply them in evaluation?” (Saracevic, 2000, p. 1).

For what purpose and goal do we evaluate?
The main purpose and goal is to determine well-designed digital libraries (WDDLs). The determined WDDLs may consist of the proposed International Open Public Digital Library (IOPDL). Also, it is, generally, to find strength and weakness of existing digital libraries. The finding may enhance their strength and complement their weakness. It will improve overall quality of digital libraries.

What should we evaluate?
We should evaluate content of collections, their interfaces’ usability and performance of existing digital libraries, because these are requirements of well-designed digital libraries. Well-Designed Digital Libraries are defined as the libraries that provide high quality content collections in a subject area(s), high quality usability and performance. WDDLs should satisfy the following requirements:
They should have their own unique high-quality collections in a special subject area(s); They should provide convenient services in usability for users to use easily them; and They should provide sufficient performance for users to access and use them without delaying time and with reliable.

These are supported by five objectives of evaluations that Saracevic (2000) divides:

“User-Centered:
Individual: How well does a digital library support information needs, tasks, and activities of people as individual users or groups of users with some strong commonalities? Interface: How well does a given interface provide and support access, searching, navigation, browsing, and interaction with a digital library?

System-centered:
Engineering: How well do hardware, networks, and related configurations perform?
Processing: How well do procedures, techniques, algorithms, and operations, and so on perform?
Content: How well are the collection or information resources selected, represented, organized, structured, and managed? ” (Saracevic, 2000, p. 14).

Who should evaluate?
The evaluation is processed at Graduate School of Library and Information Science (GSLIS) in the University of Illinois at Urbana-Champaign in 2010. The evaluation reflects many discussions and instructions of Professor McDonough, Professor Dubin, Dean Smith, and Dean Renear of GSLIS. Especially, the content evaluation was done by Professor McDonough and me. Based on instructions of the faculty, I evaluate mainly existing digital libraries.

Upon what criteria do we evaluate?
The new suggested Content, Usability and Performance Evaluation (CUPE) criteria is used, which is a set of criteria. The criteria in the set are chosen from existing evaluation criteria in order to be appropriate for the purpose of the project. It has seven criteria to evaluate content, usability, and performance.

How to apply them in evaluation? (How will the evaluation of existing digital libraries be performed?)
With the CUPE criteria, several existing evaluation methods are used, such as heuristic, using open resources, and using specially designed computer programs. To check whether existing digital libraries satisfy the requirements of well-designed digital libraries, three steps are designed to evaluate existing digital libraries:

- The first step is for experts to investigate the content of collections of existing digital libraries using a set of content evaluation sub-criteria;
- The second step is to evaluate their interfaces’ usability using open tools and heuristic methodology, and using a set of usability evaluation sub-criteria; and
- The third step is to evaluate the libraries’ performance by simulations with specially designed computer programs, using a set of performance evaluation sub-criteria.

For the first step, with ‘Content Evaluation criteria,’ digital libraries are evaluated whether they have high quality collections in a particular subject area(s). For the second step, with ‘Usability Evaluation Criteria,’ it is checked whether they provide high quality usability for users to use it without troubles and difficulties. Lastly, with ‘Performance Evaluation Criteria’ it is checked whether they provide enough performances for users to use it without delay and with reliability.

Generally, 5-point rating scale applies to evaluate the candidate digital libraries with seven criteria of CUPE criteria. Depending on each criterion, 5-point scale applies to differently.

At what level do we evaluate?
The CUPE criteria have three main Content, Usability and Performance criteria. Each Criterion has sub-criteria. Each sub-criterion has check lists, which investigate each performance in depth in varied point views.

1. The CUPE Criteria

The CUPE criteria combine three main evaluation criteria: Content, Usability and Performance evaluation criteria. Each criterion has a set of sub-criteria so that existing digital libraries can be evaluated in depth and width with the criteria. That is, the CUPE criteria have total seven evaluation criteria in three main evaluation criteria applying several methodologies. It identifies well-designed digital libraries based on their collections’ contents, usability, and performance. Such well-designed digital libraries will consist of the proposed International Open Public Digital Library.

The framework of the suggested CUPE criteria is:

- The Content Evaluation Criteria
  1) Content quality evaluation criteria
     - Accuracy - whether collections have accurate information in the subject area that the users can trust;
     - Coverage - adequacy of the scope of the collection, considering both breadth and depth;
     - Authority - how authoritative the site appears to be, based on the reputation of the organization or sponsors; and
     - Satisfaction - experts’ overall response to the digital library's collection.

- The Usability Evaluation Criteria
  2) Accessibility: whether users can access information of a digital library with no or at least limitations without errors;
  3) Interface Usability:
     3) Convenience/ ease of use: whether it is easy for users to learn and use the digital library in browsing and scanning (how much easy it is to navigate to from most pages).
     4) Interfaces’ consistency: whether it uses the same words, color, font, graphics, and layout among web pages.
     5) Visible design and Aesthetic Appeal of a digital library: whether it communicates clearly and visibly the purpose and value of the interface’s components with unique and descriptive headings and visibility, and whether it has attractive aesthetic appeal.

- The Performance Evaluation Criteria
  6) Response, retrieval time: how much time does it take to carry out tasks (navigate, browse, search, or obtain resources); the average time that a digital library takes to process all requests including the link response time and the search response time.
  7) Relevance of obtained results (effectiveness, efficiency, and usefulness): how precise obtained results are from requested queries of users.

2. The Content Evaluation Criteria

The content evaluation emphasizes that WDDLs should be representative all over the world, and their collections should be qualified and specialized uniquely in a subject area. They will consist of the proposed IOPDL for the future. Thus, the Content Evaluation Criteria should be appropriate to evaluate whether collections of a digital library are qualified and specialized uniquely in a subject area(s). To achieve the goal, faculty and I designed carefully the following particular criteria through many discussions. The content evaluation criteria include four sub-criteria.

- Accuracy - whether collections have accurate information in the subject area that the users can trust;
- Coverage - adequacy of the scope of the collection, considering both breadth and depth;
• Authority - how authoritative the site appears to be, based on the reputation of the organization or sponsors; and
• Satisfaction - experts’ overall response to the digital library's collection.

3. The Usability Evaluation Criteria

The main purpose of the suggested usability criteria is that well-designed digital libraries (WDDLS) should be accessible to users readily. For the purpose, the Usability Evaluation Criteria have two main criteria: Accessibility and Interface Usability. First, accessibility criteria are essential for the International Open Public Digital Library (IOPDL), because “Open” in the IOPDL means that anyone can access to the candidate digital libraries with no or few limitations. Thus, accessibility is one of the most important criteria to evaluate the candidate digital libraries. Also, for ease of use purpose, interfaces’ usability should be considered, because “Public” in the IOPDL means that all ages over the world can use easily WDDLS. Thus, Interface Usability Criteria have three sub criteria to evaluate their easiness: Convenience (ease of use), Interfaces’ consistency, and Visual design and Aesthetic Appeal. These criteria are chosen, because they are commonly used in the usability evaluation, and because they are suitable to evaluate many digital libraries’ usability, and related closely to the purpose of the IOPDL.

3-1. Accessibility

Accessibility is mainly chosen, because it is widely accepted as a prime element in usability evaluation (Bishop, 1998). Accessibility is defined as “the ability to receive, use, and manipulate data and operate controls included in electronic and information technology … ; whether users can access information of a digital library with no or at least limitations without errors” (IITAA). IITAA is the law for governing accessibility for State of Illinois agencies including universities. For the project, preliminary review and conformance evaluation of the below evaluation methods of websites are used to evaluate the candidate digital libraries. The detail will be discussed in chapter V.

W3C explains several accessibility evaluation methods of websites:
• “preliminary review method (to identify some potential accessibility problems);
• conformance evaluation (to determine if a Web site meets accessibility standards);
• evaluation Approaches for specific contexts (to describe evaluation during the development process, ongoing monitoring, evaluation of legacy sites, and evaluation of dynamically generated Web pages); and
• involving users in Web Accessibility Evaluation (to include people with disabilities ("users") in accessibility evaluation throughout Web development)” (W3C, Improving the Accessibility of Your Website).

3-2. Interface Usability Evaluation Criteria

The importance of Interface usability evaluation is emphasized in many papers. Nielsen defines the interface as a door to the content contained in the digital collection (Nielsen J., Usability engineering, 1993). Arms notes “(a) digital library is only as good as its interface” (Arms, 2000). As much as interfaces play important roles for digital libraries, interface usability evaluations are important for users to use them with facility. Then, which criteria can point out well problems of their interfaces of the candidate digital libraries? Because we cannot evaluate sixty three candidate digital libraries by many criteria, we should choose several important criteria. Three interface usability criteria are chosen emphasizing importance of ease of use, consistency, and visual design, because they are main factors to engage users to navigate readily websites. These three criteria will check whether a digital library can engage users without troubles or difficulties.

Convenience (Ease of Use)
As Usability.gov explains, ‘Ease of learning’ and ‘Efficiency of use’ are two of the main usability evaluation factors. Ease of use/convenience is very important for users to visit, navigate, and use digital libraries for their purposes and satisfaction. Also, all organized web pages by priority are useful to understand easily (Usability.gov). Therefore, Ease of use/Convenience evaluation criterion is chosen to check whether users can use and find information easily browsing and scanning the candidate. Convenience criterion checks: flexibility and efficiency of use; whether a digital library has help documents; whether help documents are helpful to solve errors; and whether related information and functions are clustered together. That is, it checks whether the candidate is easy for users to navigate to their most pages.

**Interfaces’ Consistency Evaluation Criteria**

In ‘Top-10 Application-Design Mistakes,” Jakob Nielsen points out that the first and second big mistakes in designing interfaces are non-standard Graphical User Interface controls and inconsistency. They cause confusion when an interface uses different words or commands for the same thing (Nielsen J., 2008). Usability.gov states guidelines to assure interfaces’ consistency: place important items consistently toward the top and center of the page; “do format common items consistently” including fonts and layout of each page; and “ensure visual consistency of website elements within and between webpages” even between data entry and data display (Usability.gov). Clearly recognizable look and feel will also engage users (Usability.gov). That is, if interfaces are consistent, it will reduce confusion for users to use them. Therefore, Interfaces’ Consistency Evaluation Criterion is chosen to check: whether interfaces are consistent among web pages; whether use the same words, situations, or actions in meaning the same thing; whether use the same color, font, graphics, and layout among web pages; and whether have consistency between data entry and data display.

**Visual Design and Aesthetic Appeal Evaluation Criteria**

The importance and guidelines of Visual design and aesthetic appeal are described in the guidelines of Usability.gov. A well-constructed (visual design) homepage will project a good first impression to all who visit the site (Usability.gov). For it, a homepage should visibly communicate the site's purpose, and contain a limited amount of text appropriately being aligned on the pages and showing a moderate amount of white space (Usability.gov). Icons should be visually and conceptually distinct but harmonious with appropriate screen density. For visual and aesthetic design, designers should use unique and descriptive headings using black text and high-contrast backgrounds colors to help users understand the grouping of related information (Usability.gov). Also, attention-attracting features such as animation, bold colors and size differentials can be used carefully. The visual design should inform about “what is going on to users in an interface … supporting undo and redo … with high search engine visibility” (Usability.gov). Therefore, Visible design and Aesthetic Appeal evaluation criterion is chosen to check whether a candidate digital library communicates clearly and visibly the purpose and value of the interface with unique and descriptive headings; whether it has visibility, informing about what is going on to users in an interface, with “emergency exit”, undo, and redo, showing path of certain directories and frequently used icons; whether it has a clearly recognizable look and feel that will engage users; whether it has minimalist design, not containing irrelevant or rarely needed information; and whether it has attractive aesthetic appeal features such as animation, bold colors and size differentials with appropriate screen density.

4. **The Performance Evaluation Criteria**

Well-designed digital libraries which will consist of the IOPDL should provide rapid and accurate responses, especially retrieval time for queries. Relevance of obtained results is also significant performance of well-designed digital libraries. Generally, they are very important performances for users to use a digital library without waiting for long time and to have confidence in obtained results. Users cannot wait for even little time for a response from their requests. Response time is one of important measurements to evaluate digital libraries. Moreover, many unreliable sources in obtained results hinder
work efficiency of users. Relevance is necessary that well-designed digital libraries should provide. Hence, two main performance criteria (response and retrieval time, and relevance of obtained result) are selected, since they are effective for performance of digital libraries. Another reason why they are chosen is they can be measured automatically by programs with Python computer language.

**Response and Retrieval Time**
Response, retrieval time is how much time it takes to carry out tasks such as navigation, browse, search, or obtain resources. Reeves (2003) emphasizes information retrieval is important in evaluating a digital library, because it reveals how effectively the information is located in a library (Reeves & Woo, 2003). Reeves also points out that retrieval information in a single digital library collection is very complex, because it involves concepts such as metadata, cataloguing and indexing. This complexity is increased if a digital library provides to search across multiple collections using different metadata systems for indexing (Reeves & Woo, 2003). Nielsen (1993) argues that response times should be fast, explaining the basic response times with three limitations:

“0.1 second is about the limit for having the user feel that the system is reacting instantaneously, meaning that no special feedback is necessary except to display the result. 1.0 second is about the limit for the user's flow of thought to stay uninterrupted, even though the user will notice the delay. Normally, no special feedback is necessary during delays of more than 0.1 but less than 1.0 second, but the user does lose the feeling of operating directly on the data. 10 seconds is about the limit for keeping the user's attention focused on the dialogue. For longer delays, users will want to perform other tasks while waiting for the computer to finish, so they should be given feedback indicating when the computer expects to be done. Feedback during the delay is especially important if the response time is likely to be highly variable, since users will then not know what to expect” (Nielsen J., Usability engineering, 1993).

**Relevance**
Johnson defines relevance as how precise obtained results are from the requested queries of users (Johnson, 2008). Some researchers point out importance of relevance. Nielsen (2010) points out that bad searching including relevance reduce the performance of digital libraries (Nielsen J., 2011). Xie (2006) points out that the performance of digital library system is related to the relevance of the retrieval results and efficiency of the retrieval process (Xie, 2006).

Many methods to measure relevance performance are used such as precision, recall, fall-out, F-measure, mean average precision, discounted cumulative gain, and etc. Google uses keyword density to measure relevance. Google says “the keyword density tool is useful for helping webmasters and SEOs achieve their optimum keyword density for a set of key terms” (SEO). The seochat.com also explains

“Keyword density is important because search engines use this information to categorize a site's theme, and to determine which terms the site is relevant to. The perfect keyword density will help achieve higher search engine positions. Keyword density needs to be balanced correctly (too low and you will not get the optimum benefit, too high and your page might get flagged for ‘keyword spamming’)” (SEO Tools – Keyword Density).

Thus, in the paper, keyword density is used to measure relevance performance, as Google does. But, the method to measure relevance is dissimilar with Google, which will be explained in the chapter VI.

**IV. The Content Evaluation**

First of all, we examine whether a digital library has a unique specialized collection in one of subject areas. Those subject areas are drawn from the Library of Congress Classification, listed below. Several similar subject areas are combined, and few subject areas are deleted from the Library of Congress Classification to simplify subject areas. The simplified fifteen subject areas are manageable and appropriate for the project evaluations.
Modified fifteen subject areas Based on the Library of Congress Classification

- Philosophy, psychology, religion
- World history and history of Europe, Asia, Africa, Australia, New Zealand, ETC
- History of the Americas
- Geography, Anthropology, Recreation
- Social sciences
- Political science, law
- Education
- Music and books on music
- Arts
- Language and literature
- Science
- Medicine
- Agriculture
- Technology
- Military science

With the modified subject areas, we investigate which digital libraries have a high quality specialized collection(s). The Content Evaluation is done with the Content Evaluation Criteria that have four sub-criteria. However, many digital libraries exist. We cannot classify all of them into their subject areas in a prototype evaluation. Some digital libraries are not clear to decide their subject areas based on their collections. Thus, the evaluation is limited regionally almost in the USA and timely on 2010.

1. Methodology

The four chosen sub-criteria of the Content Evaluation Criteria are:

- Accuracy - whether collections have accurate information in the subject area that the users can trust;
- Coverage - adequacy of the scope of the collection, considering both breadth and depth;
- Authority - how authoritative the site appears to be, based on the reputation of the organization or sponsors; and
- Satisfaction - experts’ overall response to the digital library's collection.

Heuristic Method

With the criteria, Professor McDonough and I investigated which existing digital libraries are representative and have authority in their subject domain by heuristic method. In evaluating collections’ content of existing digital libraries, we put emphasis on whether each digital library satisfies accuracy, coverage, authority, and satisfaction criteria in their collections. Through the evaluation, we recommend three to seven digital libraries in each subject area of fifteen subject areas, as candidates of well-designed digital libraries.

The designed Survey Instrument

With the recommended candidate digital libraries, we designed a survey instrument so that expert librarians of each subject area can involve in the content evaluation in their expertise. We recruited participants through membership directories provided by various library associations. Participants were invited to complete the survey, which was available through the SurveyMonkey system. The separate surveys on the Web are available at

- World and America history: [http://www.surveymonkey.com/s/D8XQRPT](http://www.surveymonkey.com/s/D8XQRPT),
- Science and Technology: [http://www.surveymonkey.com/s/DLSFCMY](http://www.surveymonkey.com/s/DLSFCMY),
In the survey, participants are asked to visit three to five digital libraries, and to give a score of 5-point scales depending on how much the digital library satisfy the four sub-criteria of the Content Evaluation Criteria.

2. The Content Evaluation Result

Generally, through the content evaluation by heuristic method, three to seven digital libraries are recommended in each subject area. Total sixty three digital libraries are recommended as candidates of well-designed digital libraries. One of them, ‘William Blake Archive,’ is duplicated in Language and literature, and in Art subject areas. It is evaluated in Art subject area. Thus, total recommended digital libraries are reduced as sixty two. Sixty two recommended digital libraries will be evaluated over and over for their Usability and Performance evaluations. As a result, the recommended digital libraries in each subject area are:

- Philosophy, psychology, religion subject areas
  - (1) Chinese Philosophical Etext archive [http://sangle.web.wesleyan.edu/etext/index.html](http://sangle.web.wesleyan.edu/etext/index.html)
- World History and history of Europe (Asia, Africa, Australia, New Zeal, and ETC)
  - (8) Africa Focus: Sights and Sounds of a Continent University of Wisconsin Digital Collections [http://digicoll.library.wisc.edu/AfricaFocus/](http://digicoll.library.wisc.edu/AfricaFocus/)
- History of the America
  - (11) Documenting the American South [http://docsouth.unc.edu/](http://docsouth.unc.edu/)
- Geography
- Social sciences
  - (21) ICPSR [http://www.icpsr.umich.edu/icpsrweb/ICPSR/access/index.jsp](http://www.icpsr.umich.edu/icpsrweb/ICPSR/access/index.jsp)
- Political science, law
(22) THOMAS http://thomas.loc.gov/
(23) GPO Access http://www.gpoaccess.gov/index.html
(24) Civil Rights Digital Library http://crdl.usgs.edu/?Welcome&Welcome

- Education
  (27) Exploratorium Digital Library http://www.exploratorium.edu/educate/dl.html
  (28) HistoryMakers http://www.thehistorymakers.com/
  (29) SMETE digital library http://www.smete.org/smete/

- Music and books on music
  (30) Database of Recorded American Music http://www.dramonline.org
  (31) Sheet Music Consortium http://digital.library.ucla.edu/sheetmusic/
  (33) E. Azalia Hackley Collection http://www.thehackley.org

- Arts
  (34) NYPL Digital Gallery http://digitalgallery.nypl.org/nypldigital/index.cfm
  (36) Art History Resources on the Web http://witcombe.sbc.edu/ARTHLinks4.html
  (37) William Blake Archive http://www.blakearchive.org/blake/

- Language and literature
  (38) ACLA (American Comparative Literature Association) http://www.acla.org/
  (41) Southeast Asia Digital Library (SADL) http://sea.lib.niu.edu/reslanguage.html

- Science
  (42) National science Digital Library http://nsdl.org/
  (44) National Science Foundation http://www.nsf.gov/
  (47) Public Library of Science http://www.plosbiology.org/home.action

- Medicine
  (49) AMA (American Medical Association) http://www.ama-assn.org/
  (50) ecancermedicalscience http://www.ecancermedicalscience.com/
  (51) U.S. Department of Health & Human Services HHS.gov http://www.hhs.gov/
  (52) Children’s Medical Center http://www.childrens.com/HealthLibrary/HealthLibContent.cfm

- Agriculture
  (54) Core Historical Literature of Agriculture in Cornell University http://chla.library.cornell.edu/
  (55) Western Waters Digital Library http://www.westernwaters.org/

- Technology
  (57) Dspace@MIT http://dspace.mit.edu/
  (59) IEEE computer society http://www.computer.org/portal/web/csdl
Results of the Survey and Analysis

With the candidate digital libraries, we would like to evaluate again their content quality of collections with the designed survey instrument. However, the survey couldn’t get much attention to expert librarians. Because of lacking of recognition how important the evaluation is, they had not invested their enough time to do the survey. Also, they were difficult to evaluate three to five digital libraries in detail. Merged subject areas were also one of reasons that made experts difficult to respond. Susan Searing who was the librarian and associate professor of GSLIS pointed out that the time required to complete the survey, 20-40 minutes, is a lot of time for a busy person to invest without any direct reward. Although the survey methods for evaluating content quality of collections may be appropriate, we cannot give credits for the poor response results. Therefore, only the evaluation result by heuristic method is accepted as the content evaluation with the Content Evaluation Criteria.

3. The Analyses and Discussion of the Content Evaluation

Many existing digital libraries are investigated in fifteen subject areas, regarding to their collections’ accuracy, authority, scope, and overall satisfaction. Through the content evaluation by heuristic method, some important facts are discovered. Overall, national digital libraries in the U.S. show very high quality and quantitative collections in accuracy, authority, breadth and depth scope, and satisfaction. Although many digital libraries exist, some digital libraries have not uniquely specialized collections in any subject areas. Some digital libraries cannot prove enough quality or quantitative collections in a subject area(s). Especially, private digital libraries seem not to have enough quantitative collections. Nevertheless, few private digital libraries (e.g., IEEE, etc.) show high quality and quantitative content collections in their domains. But, they are limited for evaluators to access by authority and fee limitations. That is, the biggest challenge in the content quality evaluation is limitations in accessing to digital libraries and their collections.

Generally, the content quality evaluation of existing digital libraries was efficient to discover high quality collections in fifteen subject areas for a short time. We found effectively and intensively the candidate digital libraries with four sub-criteria. The four sub-criteria (accuracy, authority, scope, and overall satisfaction) are appropriate to evaluate overall quality of collections. Further, the content evaluation in the paper may present a certain way to evaluate content quality of existing national digital libraries around the world. But, there are few shortages in the content evaluation of this prototype. We may not find all quality digital libraries in the U.S. There must be better digital libraries in fifteen subject areas that were not included in the candidate digital libraries. Especially, private digital libraries that have small collections tend to be excluded in the recommended digital libraries. Some of them are excluded by the limitation that blocks for evaluators to access. Also, if more experts involved in evaluating content quality in their subject domains, and if many participants willingly attended to the designed survey tool, more detailed investigations could have been done, and more hidden good digital libraries might have been found. For the future content evaluations, above all, participants should recognize how important the content evaluation is. The evaluation results can demonstrate which digital libraries can be representative digital libraries in a subject domain(s). The results of the content evaluation may point out and address some problems of existing digital libraries in their content/collections. Through the content evaluation results, existing digital libraries can be encouraged to develop their unique specialized collections further.
It will be good for the digital libraries to improve their quality generally, and for users to take the advantages from them.

V. The Usability Evaluation

By the content quality evaluation, sixty two digital libraries are recommended as candidates of well-designed digital libraries. The candidates are evaluated again by Usability evaluations\(^2\) with the suggested the Usability Evaluation Criteria:

- **Accessibility**: whether users can access information of a digital library with no or at least limitations without errors;
- **Interface Usability**:
  - **Convenience/ease of use**: whether it is easy for users to learn and use the digital library in browsing and scanning (how much easy it is to navigate to from most pages).
  - Interfaces’ consistency: whether it uses the same words, color, font, graphics, and layout among web pages.
  - Visible design and Aesthetic Appeal of a digital library: whether it communicates clearly and visibly the purpose and value of the interface’s components with unique and descriptive headings and visibility, and whether it has attractive aesthetic appeal.

With the criteria, in the chapter, first of all, methodology of Accessibility evaluation is explained. The evaluation results are analyzed and discussed. Second, methodology of interface usability evaluation is explained along with check lists and methods for each sub-criteria. Evaluation results are explained, analyzed and discussed.

1. Accessibility Evaluation

1-1. Methodology

Accessibility evaluation in Usability Evaluation Criteria investigates limitations and error degrees in accessing a digital library. Several methodologies have been presented to evaluate accessibility. For example, Giorgio Brajnik (2008) lists some possible methods of evaluating accessibility: expert review, user testing, subjective evaluations, and barrier walkthrough (Brajnik, 2008). On the other hand, W3C Web Accessibility initiative provides many tools that have been developed to evaluate accessibility automatically. I decide to use open accessibility tools, because it is simple but accurate and economical to evaluate many digital libraries in a short time. I chose seven web accessibility evaluation tools based on standards that they use and unique characteristics.

The selected seven accessibility evaluation tools use different evaluation methods based on different standards. Each accessibility evaluation tool complies with one of Illinois Information Technology Accessibility Act (IITAA), Electronic and Information Technology Accessibility Standards (Section 508), W3C Web Content Accessibility Guidelines (WCAG), etc. The Table 1 summarizes unique characteristics of the selected seven web accessibility evaluation tools. With the chosen tools, the web pages of each digital library are evaluated whether they have accessibility errors or broken links. Each digital library’s web address is inputted in each Accessibility evaluation tool. Each Accessibility evaluation tool evaluates the library’s accessibility, and gives number of errors based on its criterion. Depending on number of errors, the digital library is scored with 5-point scale. General accessibility evaluation process is:

1) Put each website address of the candidate digital libraries on each evaluation tool;
2) Each tool evaluates their accessibility, and gives a report how many accessibility errors come out, and whether it passes guidelines (e.g. Section 508, IITAA, or WCAG);

---

\(^2\) The Usability Evaluation was completed from approximately March 2010 to October 2010.
3) Analyze the result, and give scores based on 5-point scale: 5-point rating scale is used variously depending on evaluation tools and number of detected errors.

4) Calculate an average for each candidate digital library from seven scores of seven evaluation tools. The average for each candidate means a degree of accessibility through accessibility evaluations. The high average means high accessibility of a candidate digital library.

<table>
<thead>
<tr>
<th>Accessibility Evaluation Tools</th>
<th>What is the tool? Which standards does it use?</th>
<th>Why is it chosen for accessibility evaluations?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional Accessibility Evaluator (FAE)</td>
<td>FAE is developed to test web sites for functional accessibility features defined in the iCITA HTML Best Practices. The iCITA HTML Best Practices is a statement of techniques for implementation of the W3C Web Content Accessibility Guidelines (WCAG), the United States Federal Government Electronic and Information Technology Accessibility Standards (Section 508) and the Illinois Information Technology Accessibility Act (IITAA).</td>
<td>It evaluates Accessibility of a digital library by the W3C Web Content Accessibility Guidelines (WCAG) and the United States Federal Government Electronic and Information Technology Accessibility Standards (Section 508).</td>
</tr>
<tr>
<td>The HiSoftware Cynthia Says portal</td>
<td>The HiSoftware Cynthia Says portal is to identify errors in a home page that is related to Section 508 standards and/or the WCAG guidelines.</td>
<td>It checks verification checklist, according to 508 Standards section 1194.22 and/or the WCAG guidelines.</td>
</tr>
<tr>
<td>W3C Markup Validation Service</td>
<td>W3C Markup Validation Service checks the markup validity of Web documents in HTML, XHTML, SMIL, MathML, etc. It validates specific content such as RSS/Atom feeds or CSS style sheets, MobileOKcontent, or broken links.</td>
<td>It finds errors and warnings checking the web site as XHTML 1.0 Transitional. It is a unique tool that checks the markup validity of HTML and others with CSS.</td>
</tr>
<tr>
<td>Web Accessibility Evaluation Tool (WAVE)</td>
<td>WAVE tool shows the accessibility evaluation results with embedded icons and indicators in the original digital library homepage including detected error numbers. The embedded icons and indicators tell accessibility of the page.</td>
<td>It is very simple to evaluate Accessibility (their used standards are not clearly stated).</td>
</tr>
<tr>
<td>The etre Accessibility Check</td>
<td>The etre Accessibility Check checks accessibility of web sites by the WAI accessibility guidelines.</td>
<td>It checks accessibility of web sites by the WAI accessibility guidelines. The results show accessibility issues by WAI Priority.</td>
</tr>
<tr>
<td>Web Accessibility Evaluator in a single XSLT file (WAEX)</td>
<td>WAEX evaluates web accessibility by WCAG Checkpoints priority 1, 2, and 3 and W3C Recommendations.</td>
<td>It evaluates web accessibility by WCAG Checkpoints priority 1, 2, and 3 and W3C recommendations. The report shows whether the digital library passes WCAG Checkpoints of priority 1, 2, 3, and whether it is in WCAG or not.</td>
</tr>
<tr>
<td>Fujitsu Web Accessibility Inspector 5.11</td>
<td>Fujitsu Web Accessibility Inspector 5.11 checks accessibility of web sites with CSS files by WCAG Checkpoints priority 1, 2, and 3, and W3C Recommendations.</td>
<td>It checks accessibility of web sites with CSS files by WCAG Checkpoints priority 1, 2, and 3, and W3C recommendations.</td>
</tr>
</tbody>
</table>

Table 1: The selected evaluation tools
1-2. Results and Analyses
In the section, I explain scoring methods and results according to characteristics of each evaluation tool. The results of accessibility evaluations are somewhat different based on different evaluation tools. Thus, diverse analysis methods are applied to each evaluation tool and its evaluation results.

Functional Accessibility Evaluator (FAE)
Functional Accessibility Evaluator 1.0.3 evaluates Accessibility of digital libraries based on the W3C Web Content Accessibility Guidelines (WCAG) (UIUC). With urls, the digital libraries are evaluated by the FAE tool. The evaluation results of using FAE shows whether the web pages pass, fail or have warning based on WCAG guidelines. The result includes the details of the accessibility features and problems in a website. Also, the result shows how many ‘Complete,’ ‘Partially Implemented,’ and ‘Almost Complete’ occur. Based on the number of ‘Complete,’ the digital library is scored with 5-point scale, such as: 5 points: 5 ‘Complete’ in five categories, 4 points: 4 ‘Complete’, 3 points: 3 ‘Complete’, 2 points: 2 ‘Complete’, 1 point: 1 ‘Complete’, 0 point: 0 ‘Complete.’ According to number of ‘Almost Complete,’ 0.5 points are added.

For example, Figure 1 shows the evaluation result of United States Department of Defense digital library, http://www.defense.gov/, in Military science subject area. The status of five categories is shown by Complete, Almost Complete, and Partially Implemented. In the paper, only numbers of Complete and Almost Complete are counted in order to be scored with 5-point scales. Thus, United States Department of Defense digital library is 2.5 score by 2 complete and 1 almost complete.

<table>
<thead>
<tr>
<th>Category</th>
<th>Status</th>
<th>% Pass</th>
<th>% Warn</th>
<th>% Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navigation &amp; Orientation</td>
<td>Partially Implemented</td>
<td>69</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>Text Equivalents</td>
<td>Complete</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Scripting</td>
<td>Complete</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Styling</td>
<td>Partially Implemented</td>
<td>83</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>HTML Standards</td>
<td>Almost Complete</td>
<td>50</td>
<td>50</td>
<td>0</td>
</tr>
</tbody>
</table>

(Note: % Pass includes N/A results.)

<table>
<thead>
<tr>
<th>Value</th>
<th>Percent</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>100</td>
<td>Pass + N/A</td>
</tr>
<tr>
<td>Almost Complete</td>
<td>95–100</td>
<td>Pass + N/A + Warn</td>
</tr>
<tr>
<td>Partially Implemented</td>
<td>40–94</td>
<td>Pass + N/A + Warn</td>
</tr>
<tr>
<td>Not Implemented</td>
<td>0–39</td>
<td>Pass + N/A + Warn</td>
</tr>
</tbody>
</table>

Status Value Definitions

Figure 1: The Result of United States Department of Defense by FAE

Total, 4.8% digital libraries of sixty two digital libraries gained 4.5 points. The overall average of the results with FAE tool is 2.6123. This may mean that many digital libraries do not follow more than 95% of the W3C Web Content Accessibility Guidelines (WCAG) in inclusion of structural markup, using properly of images for interoperability, using of CSS styling, and supporting for HTML standards. Moreover, four digital libraries of sixty two digital libraries could not be evaluated, because of broken links or accessing errors. As a result, the tool shows the shortest time to execute evaluating accessibility among the chosen seven accessibility tools.
‘Cynthia Says’ Tool
Cynthia Says tool checks verification checklist, according to 508 Standards section 1194.22. Two of verification checklists are ‘A text equivalent for every non-text element shall be provided,’ and ‘Equivalent alternatives for any multimedia presentation shall be synchronized with the presentation’ (CybthiaSays). The results express whether web sites of the digital library pass the checklist or not. The number of ‘Yes’ for the passed checklist is counted for rating based on 5-point scales. Each ‘yes’ adds 1 point. To illustrate, ‘3 yes’ becomes 3 points out of 5 points.

Figure 2 shows a part of the results by Standards Section 1194.22, for the same digital library, United States Department of Defense, http://www.defense.gov/, in Military science subject area. The result shows one ‘Yes,’ thus, it gets 1 point out of 5-points.

<table>
<thead>
<tr>
<th>Standards, Section 1194.22, (m)</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule: 6.3.5 - All OBJECT elements are required to contain element content.</td>
<td>0</td>
</tr>
<tr>
<td>o No OBJECT elements found in document body.</td>
<td></td>
</tr>
<tr>
<td>Rule: 6.3.6 - All APPLET elements are required to contain both element content and the alt attribute.</td>
<td>0</td>
</tr>
<tr>
<td>o No APPLET elements found in document body.</td>
<td></td>
</tr>
<tr>
<td>Rule: 6.3.7 - When EMBED Elements are used, the NOEMBED element is required in the document.</td>
<td>0</td>
</tr>
<tr>
<td>o No EMBED elements found in document body.</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: The Result of United States Department of Defense by Cynthia Says

Totally, only five following digital libraries had three ‘Yes,’ because it is difficult to satisfy all standards of Section 1194.22. That is, 7.9% of the candidate digital libraries gained 3 points.

- The National Archives Online Exhibits in History of the America Subject area,
- SMETE digital library in Education Subject area,
- California Sheet Music in Music Subject area,
- U.S. Department of Health & Human Services in Medicine, and
- The National Archives Education Resources, UK, in Social Science Subject Area.

The overall average is as low as 1.580357 of seven accessibility tools. It may mean that home pages of many digital libraries could not satisfy many rules of Section 508 standards, and/or the WCAG guidelines. On the other hand, the method seems very strict to evaluate Accessibility, since it does not give degrees of percentage but give only number of ‘Yes’ when the digital library satisfies all rules. Three digital libraries could not be evaluated, because of HTTP Transfer Errors.

W3C Markup Validation Service Tool
The W3C Markup Validation Service tool finds errors and warnings checking the web site by XHTML 1.0 Transitional (W3C, W3C Markup Validation Service). The tool reports what kinds of errors are detected, and how many errors and warnings are found. Depending on the number of errors, they are scored with 5-point scales, such as 5 points: if 0-3 errors detected in the result, 4 points: if 4-7 errors detected, 3 points: 8-11 errors, 2 points: 12-15 errors, 1 point: 16-20 errors, 0 point: more than 20 errors. The digital library that has many errors gets lower score in 5-point scales, but number of warnings is ignored.
Totally, 14.3 % of the candidate digital libraries gained 5 points, but 50.8% of them gained 0 point by W3C Markup Validation Service Tool. Interestingly, more than half digital libraries of sixty two digital libraries had more than twenty errors in Markup validation. Overall average is as low as 1.5864071. Against expectations, there are many errors in codes of HTML. That is, considerable errors are detected in the markup validity of Web documents in HTML, XHTML, SMIL, MathML, etc., although the candidate digital libraries are very good digital libraries in their subject areas. It points out that digital libraries should check carefully the markup validity of HTML or XHTML codes. The tool detects many errors than other six Accessibility tools with explanations about errors in each line and column.

**WAVE (Web Accessibility Evaluation) Tool**
The Web Accessibility Evaluation (WAVE) tool shows the accessibility evaluation results of the digital library including number of detected errors with embedded icons and indicators (WAVE). The embedded icons and indicators tell accessibility of the webpage. Depending on number of detected accessibility errors, they are scored with 5-point scale such as 5 points: if 0-2 errors are detected, 4 points: 3-5 errors, 3 points: 6-8 errors, 2 points: 9-11 errors, 1 point: 12-14 errors, 0 point: if more than 14 errors are detected. The digital library that has many detected errors has lower scores in 5-point scales.

Totally, 49% of the candidate digital libraries gained 5 points. It is a considerable number. 12.7% of all digital libraries gained 0 point. The total average of accessibility results by WAVE is as high as 3.618304. Two digital libraries cannot be evaluated by WAVE Accessibility tool. Relatively, it detects fewer errors than other tools for all candidate digital libraries. While it is simple to be used, it seems not being able to detect many accessibility errors with standards.

**etre Accessibility Check Service Tool**
The etre Accessibility Check Service Tool checks accessibility of web sites by the WAI accessibility guidelines (Etre). The results of etre Accessibility Check tool state accessibility issues by WAI Priority. The result shows number of errors that the tested page does not adhere to the WAI accessibility guidelines. The Priority 1 section reports number of errors that must be fixed. The Priority section 2 reports number of errors that should be fixed. Depending on number of errors or problems of Priority 1 and 2, the result is scored into 5-point scales. The number of errors of Priority1 is scored with 5-point scale like the below left. Next, number of errors of Priority2 is scored like the below right. Then, an average is calculated with two score values of Priority 1 and 2. The averages become final accessibility evaluation scores for the candidate digital libraries by etre Accessibility Check tool.

<table>
<thead>
<tr>
<th>Based on number of Priority1 errors:</th>
<th>Based on Priority2 errors that should be fixed:</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 points: 0-1 number of errors</td>
<td>5 points: 0-2 number of errors</td>
</tr>
<tr>
<td>4: 2-3 number of errors</td>
<td>4: 3-5 number of errors</td>
</tr>
<tr>
<td>3: 4-5 number of errors</td>
<td>3: 6-8 number of errors</td>
</tr>
<tr>
<td>2: 6-7 number of errors</td>
<td>2: 9-11 number of errors</td>
</tr>
<tr>
<td>1: 8-9 number of errors</td>
<td>1: 12-14 number of errors</td>
</tr>
<tr>
<td>0: more than 10 errors</td>
<td>0: more than 15 errors</td>
</tr>
</tbody>
</table>

Totally, etre Accessibility Check tool detects few errors of Priority1 (must fix) and Priority2 (should fix) than other six accessibility tools. Thus, it has the highest average, 3.6781, of all sixty two digital libraries. 30% of digital libraries gained 5 points. It is somewhat high rate. There are no digital libraries that get 0 point, although etre Accessibility Check tool evaluates by WAI Priority1 and Priority 2. Moreover, there are no digital libraries that are detected more than 10 errors in Priority1, and that are detected more than 15 errors in priority 2. It may imply whether many digital libraries follow WAI Priority 1 and Priority 2, or whether the tool could not detect many accessibility errors than other tools.

**Web Accessibility Evaluator in a single XSLT file (WAEX)**
Web Accessibility Evaluator in a single XSLT file (WAEX) evaluates web accessibility by WCAG Checkpoints priority 1, 2, and 3 and W3C Recommendations as FAE does (WAEX). The result shows whether the digital library pass WCAG Checkpoints priority 1, 2, 3, and Checkpoints in WCAG or not. The number of ‘passed’ is counted in order to be scored with 5-point scale such as 4 points: pass 4, 3 points: pass 3, 2 points: pass 2, 1 point: pass 1, 0 point: pass 0. And 1 point is added when the digital library has no error in priority1.

Totally, the average of sixty two digital libraries by Web Accessibility Evaluator in a single XSLT file (WAEX) is the lowest, 1.4739, while FAE tool that evaluates with same standards is 2.6123. Only 19% digital libraries of them gained 3 points out of 5 points. 30% digital libraries had 0 point. The results show that many digital libraries could not pass WCAG checkpoints of Priority 1, 2, and 3, and not in WCAG. The reasons might be that they do not follow WCAG Checkpoints priority 1, 2, and 3, and W3C Recommendations in web pages with CSS files, or that the tool is very strict to evaluate them.

**Fujitsu Web Accessibility Inspector 5.11 Tool**

Lastly, Fujitsu Web Accessibility Inspector 5.11 checks accessibility of web sites with CSS files by WCAG Checkpoints priority 1, 2, and 3, and W3C Recommendations like WAEX and FAE. It also checks Fujitsu Web Accessibility Guidelines by priority level1 (‘it checks very important items such as whether the alt attribute was set’) and level 2 (‘it checks for existence of <caption> of the table in addition to priority 1’) (Fujitsu). The number of problems of priority 1 and 2 are scored with 5-point scale like the below. Then, two score values by Priority 1 and Priority 2 are calculated into an average. The average becomes the final accessibility value for each digital library by Fujitsu tool.

In the evaluation by Priority level 1: In the evaluation by Priority level 2:
5 points: 0-1 detected errors 5 points: 0-2 detected errors
4: 2-3 detected errors 4: 3-5 detected errors
3: 4-5 detected errors 3: 6-8 detected errors
2: 6-7 detected errors 2: 9-11 detected errors
1: 8-9 detected errors 1: 12-14 detected errors

Totally, although Fujitsu Web Accessibility Inspector 5.11 detects some errors, the total average is as high as 2.9989. 14.3% gains 5 points out of 5 points. Only 0.8 % of digital libraries had 0 point. Most digital libraries gained over 2.5 points. Since several digital libraries show many errors, Fujitsu seems to detect errors well by WCAG Checkpoints priority 1, 2, and 3.

**1-3. Total Analyses with Seven Accessibility Tools**

**Exceptions of Accessibility Evaluations**

Accessibility of sixty two candidate digital libraries in fifteen subject areas was evaluated seven times by the selected seven accessibility evaluation tools. Two digital libraries, ‘Digital Past’ in the history of the America subject area and ‘Electronic Cultural Atlas Initiative’ in the social science subject area could not be evaluated by all seven accessibility evaluation tools. They may have broken links or use /robots.txt. And ‘Database of Recorded American Music digital library’ in the music subject area was evaluated by only two accessibility evaluation tools, since it has broken links, HTTP transfer errors, or inability to access in five other accessibility evaluation tools. It uses /robots.txt, too. ‘/robots.txt’ is ‘The Robots Exclusion Protocol.’ Robots.txt is ‘a text (not html) file [we] put on [our] site to tell search robots which pages you would like them not to visit” (webconfs.com). Two important considerations using /robots.txt are “robots can ignore your /robots.txt. Especially malware robots that scan the web for security vulnerabilities, and email address harvesters used by spammers will pay no attention” (About /robots.txt, 2007). As a result, seven evaluation tools couldn’t access their websites by /robots.txt. They cannot be scored nor evaluated for accessibility evaluations. Also, ten digital libraries could not be evaluated by one
or two evaluation accessibility tools. If it cannot be evaluated by one or two tools, the score becomes 0. All scores by seven tools were used to calculate an average.

**Analyses based on the Averages of each evaluation tool, and Subject Domains**

Since seven accessibility tools have different evaluation methods and standards, the results of evaluation are somewhat different. The results show which guidelines or standards a digital library follows. For example, United States Department of Defense in military science subject area gains five points by etre Accessibility Check tool that checks whether the digital library does adhere to the WAI accessibility guidelines. But the same digital library gains 0 point in W3C Markup Validation Service tool that checks the web site as XHTML 1.0 Transitional. That is, the digital library follows WAI standards, but does not follow W3C Markup rules in HTML.

The Figure 3 shows averages of accessibility evaluations based on each evaluation tool. According to the average, many digital libraries of sixty two digital libraries follow Web Accessibility Initiative (WAI) accessibility guidelines that etre Accessibility Check tool uses. Also, the results show that the digital libraries follow usually WCAG standard and W3C Recommendations as averages of Fujitsu and FAE tools show. On the other hand, many digital libraries follow less Web Content Accessibility Guidelines (WCAG) Checkpoints priority 1, 2, 3, and Checkpoints not in WCAG that WAEX uses.

Although we cannot confine that accessibility evaluation results are different among subject domains, World History and history of Europe (Asia, Africa, Australia, New Zeal, and ETC) subject domain shows that it provides the highest accessibility in all seven accessibility tools. Music and books on music domain shows the lowest accessibility overall in all seven accessibility tools.

![Figure 3: Averages of Evaluation Tools](image1)

![Figure 4: Averages of Accessibility in Subject Domains](image2)

**Analyses Based on Accessibility Evaluation Tools**

FAE is defined in the iCITA HTML Best Practices that implement the W3C Web Content Accessibility Guidelines (WCAG), the United States Federal Government Electronic and Information Technology Accessibility Standards (Section 508) and the Illinois Information Technology Accessibility Act (IITAA). The HiSoftware Cynthia Says portal detects Accessibility errors that are related to Section 508 standards.
and/or the WCAG guidelines. Two standards (WCAG and Section 508) are used in both tools, but the averages of them are some different: FAE: 2.6123 vs. Cynthia: 1.58037. It may be caused from different methods to evaluate accessibility. FAE shows percentages with three kinds of status in four categories: Navigation & Orientation, Text Equivalents, Scripting, Styling, and HTML Standards. The paper counts number of ‘Complete (100%)’ and ‘Almost Complete (95-100%)’ only including over 95% results. However, the HiSoftware Cynthia Says portal shows only ‘Yes’ in 100% or ‘No’ not in 100%. The HiSoftware Cynthia Says portal does not cover below 100%. That is, the results vary by used standards and evaluation methods.

Moreover, WAEX and Fujitsu Web Accessibility Inspector 5.11 evaluates web accessibility by WCAG Checkpoints priority 1, 2, and 3 and W3C Recommendations. Like above two accessibility tools, the tools have the same patterns in averages: WAEX is 1.4739, but Fujitsu is 2.998. WAEX shows only ‘pass (100%)’ or ‘failed (not 100%).’ However, Fujitsu Web Accessibility Inspector 5.11 shows number of errors. Depending on number of errors, the number of errors is scored with 5-point scale. Thus, Fujitsu has higher average than WAEX average. That is, using different methods to evaluate Accessibility affects the results of accessibility of each digital library. Thus, it remains a question, whether the strict method that shows only 100% is still better to evaluate Accessibility or not.

W3C Markup Validation Service checks the markup validity of Web documents in HTML, XHTML, SMIL, MathML, etc. It evaluates accessibility in somewhat different point of view. The results show that there are many errors in the markup. It needs cautions when web developers build codes with markup languages.

In conclusion, it will be better for accessibility evaluation tools to give number of errors or percentages than giving the results of only pass (yes) or fails (no). Also, it will be much better, when accessibility tools give detail explanations about errors. It will give chances for web developers to fix their errors frequently. It fits with the purpose of evaluations, for digital libraries to improve their websites.

**The Highest Score Digital libraries by the Accessibility evaluation in Each Subject Domain**
The most important result of accessibility evaluations is to find which digital libraries provide good accessibility so that users can access and retrieve information effortlessly. Finally, fifteen digital libraries in fifteen subject areas are turned out that they provide better accessibility than other digital libraries in their subject areas. And they were evaluated by all seven selected accessibility tools. That is, if a digital library could not be evaluated by even one accessibility tool, the library was excluded in the best digital libraries. The Table 2 shows digital libraries by descending series. Descending series show first the highest score digital library.

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>The highest score digital library in accessibility evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicine</td>
<td>Children’s Medical Center <a href="http://www.childrens.com/HealthLibrary/HealthLibContent.cfm">http://www.childrens.com/HealthLibrary/HealthLibContent.cfm</a></td>
</tr>
<tr>
<td>Political Science and Law</td>
<td>Harvard Law School Library Digital Collections</td>
</tr>
</tbody>
</table>

21
2. Interface Usability Evaluation

2-1. Methodology
After evaluating Accessibility, sixty two candidate digital libraries are evaluated again for their interfaces usability with three Usability evaluation criteria: Convenience, Interfaces’ Consistency, and Visible Design and Aesthetic Appeal Evaluation criteria. We designed significantly check lists for each criterion so that through the check lists, usability of each digital library can be enough evaluated by each criterion. Each criterion, thus, includes five or three evaluation check lists. Also, we use a heuristic method to evaluate those digital libraries with the check lists of three criteria. I spend some time reviewing each digital library going through several times the interface of the digital library, and then evaluate closely the digital library, inspecting whether it satisfies each check list. Each check list is scored as one point when the digital library satisfies it. Each Criterion is scored with 5-point scale. But, how much it satisfies each check list is not measured, because evaluating sixty two digital libraries was considerable work. It took a long time to evaluate all of them. Three evaluated scores of Convenience, Interfaces’ Consistency, and Visible Design and Aesthetic Appeal Evaluations are used to calculate an average. The high average tells that the digital library provides good Usability in Convenience, Interfaces’ Consistency, and Visible Design and Aesthetic Appeal.

Heuristic Evaluation Method
For interface usability evaluations, usually heuristic evaluation method is used. Nielsen (1992) defines heuristic evaluation as “a method for finding usability problems in a user interface design by having a small set of evaluators examine the interface and judge its compliance with recognized usability principles” (Nielsen J., 1992). The evaluation through heuristic method can be processed by which an individual evaluator goes through the interface several times and examines the several dialogue elements and compares them with the lists of usability principles, based on the checklists of general considered heuristics (Nielsen J., Usability engineering, 1993).

For evaluating usability by heuristic method, Nielsen (1993) created ten heuristic principles:

- “Visibility of system status: The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
Match between system and the real world: The system should speak the users' language, with words, phrases and concepts familiar to the user.

User control and freedom: Users often choose system functions by mistake and will need "emergency exit" to leave the unwanted state... Support undo and redo.

Consistency and standards: Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

Error prevention: Even better than good error messages is a careful design....

Recognition rather than recall: Make objects, actions, and options visible.

Flexibility and efficiency of use: Allow users to tailor frequent actions.

Aesthetic and minimalist design: Dialogues should not contain information which is irrelevant or rarely needed.

Help users recognize, diagnose, and recover from errors: Error messages should be expressed in plain language, precisely indicate the problem, and .. suggest a solution.

Help and documentation: Even though it is better if the system can be used without documentation, it [is] necessary to provide help and documentation” (Nielsen J., 1993).

With Nielsen’s ten heuristic principles, first, we designed carefully check lists for each criterion.

**Check Lists**

**Convenience (Ease of Use)**
- Whether it is easy for users to learn and use the digital library in browsing and scanning, or whether it is easy to navigate to from most pages.
- Whether it has flexibility and efficiency of use (how fast can a user accomplish tasks in a digital library), or whether it provides the services: hidden details, size changeable (larger text), quick links, menu, advance searching, variable languages, or being able to turn back home in anywhere and adding a comment in anywhere for structure flexibility and efficiency.
- Whether related information and functions are clustered together, or whether it has less complex structures for users to find information and to surf the structure.
- Whether it has help documents and it is helpful to solve errors (how much users can recover from these errors through help documents), and
- Whether it gives users subjective satisfaction.

**Interface’s Consistency**
- Whether it uses the same words, situations, or actions in meaning the same thing in the same place.
- Whether it uses the same color, font, graphics, and layout among web pages. And,
- Whether it has consistency between data entry and data display.

**Visible design and Aesthetic Appeal**
- Whether it communicates clearly and visibly the purpose and value of the interface with unique and descriptive headings and sub headings.
- Whether it has visibility, informing about what is going on to users in an interface, with “emergency exit”, undo, redo, back to top, more, print view, link to this page, and paths.
- Whether it has minimalist design, not containing irrelevant or rarely needed information.
- Whether it has a clearly recognizable look and feel that will engage users, not showing unfocused and untidy look, so that users are willing to use the websites without hesitation.
- Lastly, whether it has attention-attracting features, such as animation, bold colors and size differentials with appropriate screen density.

**2-2. Results and Analyses of Convenience/ Ease of Use Evaluation**
Generally, almost all digital libraries provide convenience services. Users can learn and use those digital libraries without troubles in browsing and scanning or in navigating to from most pages. Almost all digital libraries cluster related information and functions together showing less complex structures for users to find information and to surf the web sites. Most digital libraries provide flexibility and efficiency
of use providing one of the services: hidden details, size changeable (larger text), quick links, menu, advance searching, variable languages, or being able to turn back home in anywhere and adding a comment in anywhere for structure flexibility and efficiency. However, some digital libraries do not provide help documents for users to solve their problems by themselves, even though they are the candidate digital libraries in each subject domain. Lastly, few digital libraries give user satisfaction in Usability.

As a result, 36.5% of sixty two digital libraries satisfy perfectly five check lists for Convenience/ Ease of Use. 77.8% of sixty two digital libraries satisfy more than four check lists of the interface criteria. Only 6.3% of those digital libraries gained 1 or 2 points in 5-point scale, as they provide lower convenience. Therefore, generally we can say that many digital libraries of sixty two candidate digital libraries provide convenience/ease of use.

**International Children’s Digital Library Case**
For instance, International Children’s Digital Library, [http://en.childrenslibrary.org/](http://en.childrenslibrary.org/), provides very good Usability in Convenience/ Ease of Use. It is very easy to use and navigate the digital library, because it is developed for children to be able to access any information in it. In the Figure 5, the web page of International Children’s digital library shows that related information is clustered together. The library offers much flexibility and efficiency providing services: being able to turn back home page in anywhere, returning previous page, providing variable languages and sorting methods. Besides, the bottom menu improves flexibility and efficiency in use. As the Figure 6 shows the library provides good help documents in every page. Thus, we can say that it gives enough user satisfaction in Convenience/ Ease of Use.

![Figure 5: Showing clustered related information](http://en.childrenslibrary.org/)

![Figure 6: Showing Help Documents](http://en.childrenslibrary.org/)

**Special Analyses for Music Subject Domain**
The special analyses of two digital libraries in Music Subject Domain are discussed here: Database of Recorded American Music and Sheet Music Consortium. First, ‘Database of Recorded American Music’ provides very good Usability in Convenience/ Ease of Use. It is very easy to use and navigate or browse the digital library. Users can find sources by composers, performers, ensembles, instruments, and labels. But, ‘Sheet Music Consortium (SMC)’ provides a little weak ease of use. For example, Figure 7 shows the result of choosing ‘browse’ menu. Usually, when we click the ‘browse’ menu, we expect some browsed information, so that users can find information by title, authors or alphabetical order, not like searching engine. But, the result is like searching engine. When I click ‘Browse’ inside of Browse menu,
at last, the result shows ‘Browse all titles, for all dates, from all repositories, order by title.’ It might cause confusion for users to use easily the digital library.

Secondly, in the case of ‘Sheet Music consortium,’ as Figure 8 shows, it also provides clustered resources by related information, and flexibility & efficiency of use providing services: access online, more info, add, and help document. However, it cannot give user satisfaction in Convenience/ Ease of Use, because there are several confusions for users to use it. To illustrate, when I click ‘list virtual collections’ in the top menu, ‘View/Edit’ menu doesn’t work (although they are Public Access not Password Access). Moreover, there are several doubtful sources that owner names are not clear such as 123, asd, b, sd, fafasdgdgevd, etc.

As Figure 9 shows, ‘Database of Recorded American Music’ shows that related information is clustered together. When I chose ‘Instrument’ -> ‘percussion’ -> ‘bells’ -> ‘works(3)’, the result shows well clustered sources. Also, the library offers much flexibility and efficiency of use providing services: being able to turn back home page in anywhere, returning previous page. The library provides good help documents in every page, as Figure 10 shows. Thus, I can say it gives enough user satisfaction in Convenience/ Ease of Use.
2-3. Results and Analyses of Interface Consistency Evaluation
Many digital libraries provide Interface Consistency among web pages. Most digital libraries provide consistency in using the same words, situations, or actions in meaning the same thing in the same place. And most digital libraries use the same color, font, graphics, and layout among web pages. However, some digital libraries do not keep consistency between data entry and data display. Actually, it is difficult to keep consistency between data entry and data display.

As a result, over half, 58.7%, of sixty two digital libraries satisfy perfectly three check lists of interface consistency. Only 3.2% of sixty two digital libraries gained the lowest point, 2 points out of 5 points by three check lists of interface consistency. The results show that the candidate digital libraries have higher Interface consistency than Convenience and Visible design and Aesthetic Appeal in Usability Evaluation. The reason why they provide much consistency than others may be that consistency is essential in designing web sites.

**International Children’s Digital Library Case**
Figure 11 shows International Children’s Digital Library has consistency even between data entry and data display along with Figure 5 and 6. Overall, it provides very good consistency.

![Figure 11: Showing Consistency between data entry and data display](image)

Generally, four candidate digital libraries in Music Subject Domain provide Interface Consistency among web pages: Database of Recorded American Music, Sheet Music Consortium, California Sheet Music, and E. AzaliaHackley Collection. Three digital libraries use the same color, font, graphics, and layout among web pages. But, 19th-Century California Sheet Music does not use the same color, font, graphics, and layout among web pages. Also, three digital libraries except Sheet Music consortium provide consistency in using the same words, situations, or actions in meaning the same thing in the same place. In Sheet Music Consortium, meanings of ‘Current virtual collection’ and ‘list virtual collection’ menu are not clear for novice and normal users, because the word make confusion. In keeping consistency between data entry and data display, all four digital libraries do not provide the service. Actually, it seems difficult to keep consistency between data entry and data display. Even DRAM case shows also inconsistency between data entry and data display, as Figure 12 shows, although it keeps Interface Consistency among web pages.
2-4. Results and Analyses of Visible Design and Aesthetic Appeal Evaluation

Based on five check lists for Visible Design and Aesthetic Appeal criteria, the results are analyzed. Mostly, the candidate digital libraries communicate clearly and visibly the purpose and value of the interface with unique and descriptive headings and sub headings. Also, they have minimalist designs, not containing irrelevant or rarely needed information. Many digital libraries provide a clearly recognizable look and feel that will engage users, not showing unfocused and untidy look. So users are willing to use the websites without hesitation. However, some digital libraries (not many of them) provide obvious visibility. They inform about what is going on to users in an interface, with “emergency exit”, undo, redo, back to top, more, print view, link to this page, and paths. Also, few digital libraries show attention-attracting features, such as animation, bold colors and size differentials with appropriate screen density. Especially, Art and Science subject domains’ interfaces show notable Aesthetic Appeal with attention-attracting features with animation, bold colors and size differentials with appropriate screen density.

As a result, 34.9% of the candidate digital libraries satisfy perfectly the presented five facts/check lists. Only 7.9% of those digital libraries gained 2 points out of 5 points with the five factors or checklists. Although the total average of Visible Design and Aesthetic appeal criteria is lower than the total average of Interface Consistency criteria, many candidate digital libraries seem to provide good Visible Design, but less Aesthetic appeal in Usability Evaluation.

Analyses Based on Subject Domains, and International Children’s Digital Library Case

Overall, language and literature subject domain has the highest average value in Visible Design and Aesthetic appeal. On the other hand, education subject domain has the lowest average value in Visible Design and Aesthetic appeal evaluation. For instance, International Children’s Digital Library show many visible designs with pictures, signs, and icons, informing about what is going on to users in an interface, with “emergency exit”, undo, back to home, link to this page, and paths. Also, it provides aesthetic appeal designs with a clearly recognizable look and feel that will engage users, not showing unfocused and untidy look. The digital library communicates clearly and visibly the purpose and value of the interface with unique and descriptive headings and sub headings. Since it is an international digital library for children, their design is very high quality with many visible designs. It helps children access easily the web site, no matter whether they can read informed information or not.

Special Analyses for Music Subject Domain

Based on five facts/checklists for Visible Design and Aesthetic Appeal, the results are analyzed. All four chosen digital libraries in Music subject domain have minimalist designs, not containing irrelevant or
rarely needed information. Also, all four digital libraries provide a clearly recognizable look and feel that will engage users, not showing unfocused and untidy look. Three digital libraries except Sheet Music Consortium communicate clearly and visibly the purpose and value of the interface with unique and descriptive headings and sub headings. Also, those three digital libraries provide obvious visibility, informing about what is going on to users in an interface, with “emergency exit”, undo, redo, back to top, more, print view, link to this page, or paths. Sheet Music Consortium provides a little weak feeling that will engage users. But it does not show unfocused and untidy look.

Only the digital library, DRAM, however, show attention-attracting features with animation, bold colors and size differentials with appropriate screen density. To illustrate, DRAM provides many visible designs with pictures, signs, and some icons, informing about what is going on to users in an interface. It provides signs for previous page or next page. It provides main menu in any web pages. Besides, it provides aesthetic appeal designs with a clearly recognizable look and feel that will engage users. The digital library communicates clearly and visibly the purpose and value of the interface with unique and descriptive headings and sub headings. As a result, DRAM Digital Library provides Convenience ( Ease of Use) so that the novice and other users can use it without difficulties, satisfying five checklists for Convenience Usability Criteria. It provides Interface Consistency, satisfying three checklists, except showing inconsistency between data entry and data display. Also, it provides many visible designs with pictures, signs, and some icons. Lastly, it provides aesthetic appeal designs with a clearly recognizable look and feel that will engage users, not showing unfocused and untidy look. Therefore, in music subject domain, DRAM digital library can be a well-designed digital library in Usability Evaluations.

3. Analyses of Interface Usability Evaluations with Three Criteria

Differences between Typical Heuristic Evaluation and the Used Method in the Prototype

Nielsen argues that involving multiple evaluators is more possible to improve the effectiveness of heuristic method, normally three to five evaluators. An example of cost of a heuristic evaluation is $3,700 - $4,800. The adjustable cost of each evaluator is estimated to $410 -$900 (Nielsen J., 1995).

However, the prototype is to evaluate generally Interfaces’ Usability of sixty two digital libraries for several months. If we want to involve more evaluators, we should find more than twenty to thirty evaluators so that each evaluator can evaluate merely two or three digital libraries deeply and in detail. The total cost for a heuristic evaluation is estimated to $8,200- $27,000. It is huge amount for a graduate student to pay without funding. Therefore, the heuristic evaluation for the project is a little different with the general heuristic evaluation method that Nielsen describes. That is, the project places emphasis on investigating whether the digital libraries provide general Interfaces’ Usability with the carefully prepared checklists. I evaluated sixty two candidate digital libraries for several months. I investigate whether the digital libraries provide each check list for each criterion, not being able to investigate the interface minutely, and to detect detail usability problems deeply with detail results. But, the check lists for each usability criteria prepared thoughtfully so that various viewpoints of usability for each usability criteria can be evaluated. I spent some time to investigate and navigate each digital library, navigating most web pages for each digital library. As a result, the checklists for each Usability criteria play crucial roles to evaluate digital libraries in each subject area.

Analyses based on Averages of Three Interface Usability Evaluations

The Figure 13 shows averages of Convenience, Interfaces’ Consistency, and Visible design and Aesthetic Appeal Usability Evaluations in fifteen subject areas. While some digital libraries show similar averages among evaluations by three usability criteria, several digital libraries such as in Education subject domain show wide differences. This may infer that some digital libraries should improve especially some fields of Convenience, Consistency of interfaces, or Visible design and Aesthetic Appeal.
The Figure 14 shows comparisons of total averages among fifteen subject domains, which are calculated by three averages of three usability evaluations. The digital libraries in Language, Art, and Geography subject domains show the highest average in Interface Usability evaluations. But, the digital libraries in Philosophy, Education, and Military Science subject domains show the lowest average in Interface Usability evaluations with three interface usability criteria: Convenience, Interfaces’ consistency, and Visible design and Aesthetic Appeal.

**Total Analyses and discussions of Interface Usability Evaluations**

To sum up the results of evaluating Usability, many digital libraries provide Convenience services clustering related information together, and Interface Consistency among web pages. But, oddly, some digital libraries still do not provide help documents for users to solve problems by themselves, although they are candidates of well-designed digital libraries in fifteen subject areas. Providing flexibility and
efficiency of use needs to be developed in existing digital libraries. Some digital libraries do not inform about what is going on to users in an interface. Thus, providing more visibility is required to be developed. Consistency between data entry and data display turns out as the weakest point in Interfaces’ Usability of the digital libraries. Although it is difficult, it may be worthy to keep consistency between data entry and data display. Lastly, very few digital libraries provide attention-attracting features in Aesthetic Appeal Usability evaluation. However, it is true that users are willing to use attention-attracting features’ interfaces. Providing Aesthetic appeal is essential to engage more users. The detail scores are in Appendix A along with other evaluated scores.

*The highest Average digital library by Interface Usability Evaluations in each subject domain*

The Table 3 shows the digital libraries that show the highest average in each subject domain by a descending series. The evaluations are implemented with three usability criteria: Convenience, Interface Consistency, and Visible Design and Aesthetic appeal criteria.

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>The Highest Score Digital Library(ies) in each Subject Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;Perfect Score: 5&gt;</td>
<td>(5: Total average of three evaluation scores)</td>
</tr>
<tr>
<td>Political Science and Law</td>
<td>Civil Rights Digital Library <a href="http://crdl.usg.edu/?Welcome&amp;Welcome">http://crdl.usg.edu/?Welcome&amp;Welcome</a></td>
</tr>
<tr>
<td></td>
<td>Southeast Asia Digital Library (SADL) <a href="http://sea.lib.niu.edu/reslanguage.html">http://sea.lib.niu.edu/reslanguage.html</a></td>
</tr>
<tr>
<td>Agriculture</td>
<td>Database of Recorded American Music <a href="http://www.dramonline.org">http://www.dramonline.org</a></td>
</tr>
<tr>
<td>Music and books on music</td>
<td>National Institute of Standards and Technology</td>
</tr>
</tbody>
</table>

30
Table 3: The Ranking by Interface Usability Evaluations with three Usability Evaluation Criteria

<table>
<thead>
<tr>
<th>Table 3: The Ranking by Interface Usability Evaluations with three Usability Evaluation Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
</tr>
<tr>
<td>Dspace@MIT <a href="http://dspace.mit.edu/">http://dspace.mit.edu/</a></td>
</tr>
<tr>
<td><a href="http://digicoll.library.wisc.edu/AfricaFocus/">http://digicoll.library.wisc.edu/AfricaFocus/</a></td>
</tr>
<tr>
<td>Military science</td>
</tr>
<tr>
<td><a href="http://digital-library.usma.edu/collections/">http://digital-library.usma.edu/collections/</a></td>
</tr>
<tr>
<td>United States Department of Defense <a href="http://www.defense.gov/">http://www.defense.gov/</a></td>
</tr>
<tr>
<td>Education (4)</td>
</tr>
<tr>
<td>SMETE digital library <a href="http://www.smete.org/smete/">http://www.smete.org/smete/</a></td>
</tr>
<tr>
<td>Philosophy, psychology, religion</td>
</tr>
<tr>
<td><a href="http://www.thearda.com/">http://www.thearda.com/</a></td>
</tr>
</tbody>
</table>

VI. The Performance Evaluation

The Performance Evaluation assesses Response time (retrieval time) and Relevance of obtained results of queries for the candidate digital libraries, with the Performance Evaluation Criteria:

- Response, retrieval time: how much time does it take to carry out tasks (navigate, browse, search, or obtain resources); the average time that a digital library takes to process all requests including the link response time and the search response time.
- Relevance of obtained results (effectiveness, efficiency, and usefulness): how precise obtained results are from requested queries of users.

For performance evaluations, computer programs are designed with Python computer language. Depending on each evaluation criteria, each methodology and equation are differently applied to evaluate the candidate digital libraries.

1. The Response / Retrieval Time Evaluation

1-1. Methodology

In the paper, Response, retrieval time, is defined as how much time it takes to carry out tasks such as navigation or browsing links, and/or searching or obtaining resources. It is calculated as the average time that a digital library takes to process all requests. The average response time is:

\[
\text{Average response time} = \frac{\sum_{i=1}^{n} t_i}{n}
\]

\((t_i = \text{time that a digital library takes to process a task}, \quad n = \text{all tasks number})\)

Equation 1: Average Response Time

In detail, the Response/Retrieval Time Evaluation Criteria measure the response time that takes to access to all included links in the home page (I call it as ‘the link response time’), and that takes to show results for a query in a search engine (I call it as ‘the search response time’).

The Link Response Time

First, the link response time is to measure how much time does it take to access to all linked websites in a home page of a digital library. That is, the link response time measures all accessing and responding times to all links of a home page in a digital library. Then, an average of the link response time will be calculated for each digital library of the candidate digital libraries.

---

3 The performance evaluation was completed approximately from October 2010 to December 2010.
Average Link response time = \frac{\sum_{i=1}^{n} lti}{n}

(lti = time that a digital library takes to access and response to a link of a home page
\quad n = Number of all links in a home page)

Equation 2: Average Link Response Time

The Search Response Time
The search response time measures the response time of search engines to queries. The search response time measures how long it takes to upload search results, when a query is inputted in its search engine of a digital library. That is, it measures how much time a search engine takes to respond and retrieve search results to queries. Since Zhang and Si point out the search response time might differ for different types of queries (Zhang & Si, 2009), I designed a program that chooses different queries from words of the home page of each digital library. It uses a standard stoplist (e.g., symbols, prepositions, and other unnecessary words). That is, each digital library will have different queries depending on its subject domain and its content. And, the search response time is measured by the response time for all queries (all words except stoplist) in a home page of each digital library. Then, an average of the search response time will be calculated for each candidate digital library with the below equation.

Average search response time = \frac{\sum_{i=1}^{n} sti}{n}

(sti = time that a digital library takes to upload/respond search results for a query (word) in a home page, n = Number of all queries in a home page)

Equation 3: Average Search Response Time

Outline of the Response Time Evaluations
With the link and search response time criteria, the outline of the response time evaluations is:

1) An average link response time is measured with accessing times to all links in the source code of the home page of each digital library by the Equation 2.

2) Next, averages of the search response time will be calculated for each digital library of sixty two digital libraries. The search response time of search engines measures how much time it takes to upload the search results for queries with the Equation 3. That is, the search response time measures the response time for queries. The queries are created from all words in title <title> and paragraphs <p> in the source code of the home page of each digital library, filtering out common stop words (e.g., symbols, prepositions, and other unnecessary words).

3) Lastly, two averages of the link and search response time will be calculated as an average response time for a digital library, because it should be one value for a response (retrieve) time evaluation.

An Average response time for Response time evaluation

\frac{(an \text{ average link response time} + an \text{ average search response time})}{2}

Equation 4: An Average Response Time for Response time evaluation

The Detail Procedure with the Designed Computer Program
To implement the methodology for evaluating the response time, a program is designed with Python Programming Language. The program measures automatically the link response time and the search response time. To measure the link response time, the designed program does:

- First, the program finds all links from the source code of each digital library’s home page.
- Next, it measures the time, how long it takes to access to the linked websites. The below shows a part of the program that measures the linked response time.
t0=time.time()  #initialize the initial time
try:
    page=urllib.urlopen(dicitem)  #open a website of ‘dicitem’ dictionary.
    except IOError as (errno, strerror):  #if it meets an error of except IOError, continue #continue/skip
    pageload=page.read()  #read the website
    t1=time.time()  #after finishing uploading the website, measure the finishing time.
    linkresponsetime=linkresponsetime+(t1-t0)  #to accumulate the link response time, add # the response time, (t1-t0) into linkresponsetime.

● Lastly, the link response times of all links are summed up as a total time. The total time is divided by the number of all links, to calculate an average for each digital library:
    averageresponsetime= linkresponsetime/float(number).

To measure the search response time:
● First, the program makes queries for the search engine from words of <title> and <p> in the source code in the home page of each digital library. The words in <title> and <p> are investigated whether they are prepositions, signs, or meaningless words to filter out common stoplist. If they are in the common stoplist, they are not included into the query dictionary. Except them, the other words are saved into the query dictionary as queries.
● Next, to measure the search response time, each query in the query dictionary is inputted into the search engine of each digital library. To input queries into a search engine, search engines of digital libraries are investigated: how digital libraries implement their search engines, and how queries of the query dictionary can be inputted automatically into the search engine. Based on search engines, the program for search response time is designed.
● Then, the response time that takes to upload the searched results is measured for all queries.
● An average of the search response time is finally calculated.

Next, an average of response time evaluations for a digital library is calculated by the average of the link response time and the average of the search response time. Finally, the average is scored with 5-point scale method as other evaluations have been done. Depending on the average response times, the used score method is as 5 points: from 0 to less than 0.3 second’s average response time, 4 points: from 0.3 to less than 0.6 second, 3 points: from 0.6 to less than 0.9 second, 2 points: from 0.9 to less than 2.0 second, 1 point: over 2.0 second, 0 point: in the case of that a digital library cannot show the response time at all.

Prerequisite
Before the experiment for the response time evaluations, the impact on locations of digital libraries was seriously discussed. If the location of a digital library is nearby Champaign and Urbana where the experiment is done, the response time must be shorter than far distance digital libraries. Thus, to reduce the impact on locations, the experiment was done several times with the server of Graduate School of Library and Information Science not using personal laptops or desktops. Also, considering time impact online, the experiment was done several times with various times.

1-2. The Result of the Response Time Evaluation

Examples of the Results
This section shows the results of evaluating the link and search response times. They are implemented by the specially designed program with Python program language. For example, the following is a part of the result of the link response time of American Journeys digital library, http://www.americanjourneys.org/.

t0=1288639003.05  #initialize time for each link
t1=1288639003.12  #the time after accessing the link
linkresponsetime= 0.0665850639343  #calculated link response time
Second, the following is a part of results to measure search response time for AMA (American Medical Association) digital library.

query=jama # JAMA is the Journal of the American Medical Association
search link: http://search0.ama-ssn.org/search/search?database=public+amnews&query=jama
st0=1290309531.15
st1=1290309533.49
searchRTtime= 2.34000015259

AMA (American Medical Association) search response Time: [0.43799998079027447, 1.0636029770001585]

Lastly, the averages of the link response time and of search response time are calculated by the program, like the following:

American Journeys average RT1= 0.811272897077
Public Library of Science average RT1= 1.82380480976
ICPSR averageRT1= 0.623607204074

General Results
As a result, Southeast Asia Digital Library (SADL) digital library shows the shortest response time in link and search response time as 0.023540511727333069 second. 'Chinese Philosophical Etext archive' digital library shows 0.11893844646350098 response time. NASAs Visible Earth shows 0.1227287252744039 response time, Military History and Military Science of The Library of Congress shows 0.22897380122951433 second. U.S. Department of Health & Human Services HHS.gov shows 0.23830916853654305 second and so on. It looks obvious that NASAs and Military digital libraries under the U.S. government show the fastest response times.

Generally, about 74% of sixty two candidate digital libraries are evaluated by both link and search response time evaluations. About 18% of them showed only link response time, not being able to access search engines or not being able to input queries in the offered search engine. About 0.05% of them could not be evaluated by the program, because they cannot be accessible by robots.txt or by other reasons. Digital Past digital library cannot be evaluated at all, because it has broken links. Also, SMETE digital library, Database of Recorded American Music, and Analytical sciences Digital Library (ASDL) did not show any response times.

1-3. The Analyses of the Response Time Evaluation
As I describe in prerequisite, before the experiment for the response time evaluation, the impact on locations and time zones are sincerely discussed. To reduce the impact on locations and time zones, the experiment was done several times with the server of Graduate School of Library and Information Science. However, the result is somewhat different than we expected. The Southeast Asia Digital Library (SADL) and Chinese Philosophical Etext archive digital libraries show the shortest response time no
matter how much their digital libraries are far from Champaign in Illinois in the U.S. Therefore, I can tell
that the distance of the digital libraries doesn’t effect on online response times.

The Analysis based on Scores of Each Digital library
According to the Figure 15 chart, 24% digital libraries of sixty two digital libraries get 3 points out of 5
points (3: a digital library shows from 0.6 to less than 0.9 second’s average response time). Totally, 66%
digital libraries show less than 0.9 second average response times.

![The Response Time Rates](image1.png)

![Averages of the Response Time in each Subject Domain](image2.png)

Figure 15: The Response Time Rates
Figure 16: Averages of Response Time

Overall, therefore, the response times of sixty two candidate digital libraries are very short. That is, we
may conclude that many candidate digital libraries provide good response time services. As Nielsen
points out, three digital libraries (Southeast Asia Digital Library (SADL), Chinese Philosophical Etext
archive, and NASAs Visible Earth) show around 0.1 second response time that “the user feel that the
system is reacting instantaneously.” Almost digital libraries provide the response time of more than 0.1
but less than 1.0 second, which “the user does lose the feeling of operating directly on the data.” Only one
digital library shows over 10 seconds response time that “is about the limit for keeping the user's attention
focused on the dialogue” (Nielsen J., Response Times: The 3 Important Limits, 1993).

The Analysis based on Subject Domains
If we analyze the results of the Response Time based on subject domains, digital libraries in Technology
and Language subject areas show the fastest response times. Then, Military Science and Geography
subject domains’ digital libraries show faster response times. It is obvious, because technology, military
science and geography subjects are closely related to speed of response time.

2. The Relevance Evaluation

2-1. Methodology
The response time is decided by an average of two measured link and search response times. The
Performance Evaluation Criteria of the CUPE criteria will consider lastly the relevance of obtained results
from the requested queries. It measures accuracy, effectiveness, efficiency, and usefulness of the search
engine. In the prototype, the relevance criteria measure how much relevant retrieved websites are to the
query. It is measured by calculating how many words are matched with the query in retrieved websites. This method is used by Google Rankings Ultimate SEO Tool as keyword density (SEO Tools – Keyword Density). That is, the relevance of the CUPE criteria bases on Keyword Density. Keyword density may be not appropriate for websites that have many images and sounds instead of words. However, keyword density is generally an efficient method to assess relevance of obtain results and how much obtained websites are related to the query. The keyword density and relevance rate are calculated like the below.

Keyword Density, \( ki \)

\[
= \frac{\text{(number of the query word in a retrieved webpage)}}{\text{(total number of words in the web page)}} \times 100
\]

\[\text{Equation 5: Keyword Density}\]

Average Relevance Rate

\[
= \frac{\sum_{i=1}^{n} ki}{n}
\]

where \( ki \) = keyword density of a webpage, \( n \) = total number of the retrieved websites

\[\text{Equation 6: Relevance Rate}\]

Outline of the Relevance evaluation

With the Relevance Rate equation, therefore, overall method to measure relevance in the paper is as follows:

- First, the queries are created from the words in the home page of each digital library, filtering out common stop words (e.g., symbols, prepositions, and other unnecessary words). The queries are relevant closely to the subject area and contents of a digital library.
- Next, each query in a digital library is inputted as a search query in its search engine.
- Then, the relevance of the query results is measured for each retrieved website by keyword density with Equation 5 and 6. Thus, the Relevance rate is as high as the sum of Keyword Density is high.

The detail with the designed computer program

To measure the relevance of search results, the program that was used to measure the search response times is modified in order to measure relevance from the retrieved results of queries. The queries that are used in measuring search response time are used to measure relevance, too. While some digital libraries have few queries that are retrieved from title and paragraphs in a home page of the digital library, some others have many queries. Also, while some digital libraries retrieve few search results for the related websites, some other digital libraries provide many retrieved websites. For example, if the program input a query, “welfare,” in the search engine of THOMAS digital library in the political science and law subject area, 189 websites are retrieved for the query.

The Relevance program, thus, measures the relevance with limitations. Only thirty retrieved websites by priority are evaluated to measure relevance, because it takes very long time to investigate all retrieved websites for all queries. But, all chosen queries are used to measure relevance for each digital library. The keyword density of each retrieved website for each query is measured with the words of title and paragraphs from the source code of the retrieved website. An average relevance rate is calculated for each candidate digital library. The below is a part of the program that measures of relevance rates.

```
worddensity=(querywordnum/float(totalwordnum))*100  #calculate a keyword density
........  #for a retrieved website
Keyworddensity= keyworddensity+worddensity  #accumulate keyword densities
........  #of retrieved websites for a query
if keyworddensity!=0 and numsites!=0:
    relevancerate=keyworddensity/float(numsites)#calculate a relevance rate for a
```
Lastly, the averages are scored with 5-point scale method like other evaluations, such as 5 points: higher than 1.5 average relevance rate, 4: between 1.0 and less than 1.5, 3: between 0.4 and less than 1.0, 2: between 0.2 and less than 0.4, 1: between 0.01 and less than 0.2, 0: less than 0.01.

2-2. The Results of Relevance Evaluation
The following shows a part of the results of the relevance program for the AMA (American Medical Association) digital library:

```
number of library: 2
item= AMA (American Medical Association) :
queries= ['american', 'medical', 'association', 'physicians', ..... , 'membership', 'benefits']
relevancelinki= http://www.ama-assn.org/ama/home/index.shtml
keyworddensity= 3.01724137931
keyworddensity= 4.42569208354
.....
AMA (American Medical Association) Relevance rates=[2.4631655742753487, 3.3851149541458274, 3.1730287433066486, 3.8922733567125487, ..... , 0.013904338153503892, 2.1531280774541477, 0.7164243810987905]
.....
```

As a result, about 13% digital libraries of the candidate sixty two digital libraries show good relevance performance. U.S. Department of Health & Human Services HHS.gov digital library shows the highest relevance in the relevance evaluations with 149 queries. On the other hand, about 35% digital libraries show the lowest relevance with queries.

2-3. The Analyses of the Relevance Evaluation
In the experiment, since we investigate sixty two digital libraries in fifteen subject domains, queries could not be settled and sampled for all subject domains and all digital libraries. To differentiate each digital library in its subject domain, all words in the title and paragraphs are investigated as queries. The words are recruited from the source codes of a home page in each digital library. Numbers of queries vary from 0 to 634. Art History Resources on the Web digital library has the most many queries, 634 queries. Thus, it took very long time to simulate the Relevance program. But, I decided that more queries would give more accurate relevance results.

**Limitations and Difficulties**
Nevertheless, the relevance evaluation has limitations. That is, two words combinations as queries are not applied to measure the relevance. Synonyms of queries also are not investigated to count matching words. The digital libraries that have merely images were difficult to get good relevance by the Relevance program, because it measures only keyword density. Science Photo Library is one of the cases.

There are some difficulties in executing the relevance program. Digital libraries do not use a unified standard form for urls for search engines. It was not simple to evaluate all sixty two digital libraries in a
program embracing many unique search urls and methods. Three digital libraries cannot be evaluated, because their search urls will not accept new queries. To illustrate, Online Medieval & Classical Library uses google search and java script without using a search url. Africa Focus: Sights and Sounds of a Continent University of Wisconsin Digital Collections uses id numbers like, 'http://diginicoll.library.wisc.edu/WebZ/SearchOrBrowse?sessionid=01-57990-112842062’, instead of using queries. The id numbers are unknown. And, Library of Congress: American History & Culture’s url is a little bit vague not being able to be used in the Relevance program. Although three digital libraries provide search engines so that users can retrieve information from their websites, they cannot be evaluated by the Relevance program. Digital Past digital library cannot be evaluated, because it has broken links.

The Analysis based on Scores of Each Digital library
According to the Relevance Rates, 13% digital libraries of sixty two digital libraries show good relevance rates (more than 1.5). But, 40% digital libraries show very low relevance rates (less than 1), including the digital libraries that cannot be evaluated by not being able to access for search engines or search urls. In the 40% group, 35% digital libraries show more than 0.4 relevance rates. As a result, overall relevance rates are much lower than other evaluation results.

The Analysis based on Each Subject Domain
Overall, the averages of relevance rates in each subject domain are low as Figure 18 shows. Geography and Military Science subject domains’ digital libraries show better relevance rates than other subject domains. Music subject domain’s digital libraries show the lowest relevance rates. But, we should recognize that they provide music sheets that do not include texts. The relevance program examined only texts. Thus, we cannot conclude that the digital libraries in the music subject domain do not provide enough good relevance by the result of the program.

3. The Ranking by the Performance Evaluations

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>The Highest Score Digital Library (ies) in each Subject Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>Digital Library/Resource</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>Agriculture (4)</td>
<td>Western Waters Digital Library <a href="http://www.westernwaters.org/B40">http://www.westernwaters.org/B40</a></td>
</tr>
<tr>
<td>Geography (3.5)</td>
<td>University of California Digital library-calisphere <a href="http://www.calisphere.universityofcalifornia.edu/">http://www.calisphere.universityofcalifornia.edu/</a></td>
</tr>
<tr>
<td>Military science (3.5)</td>
<td>US Military Academy Digital Collections <a href="http://digital-library.usma.edu/collections/">http://digital-library.usma.edu/collections/</a></td>
</tr>
<tr>
<td>Technology (3)</td>
<td>Philosophy Resources at Harvard <a href="http://hcl.harvard.edu/research/guides/philosophy/">http://hcl.harvard.edu/research/guides/philosophy/</a></td>
</tr>
<tr>
<td>Agriculture (3)</td>
<td>British Library Online Gallery <a href="http://www.bl.uk/onlinegallery/index.html">http://www.bl.uk/onlinegallery/index.html</a></td>
</tr>
<tr>
<td>Geography (3)</td>
<td>Core Historical Literature of Agriculture in Cornell University <a href="http://chla.library.cornell.edu/">http://chla.library.cornell.edu/</a></td>
</tr>
<tr>
<td>Science (3)</td>
<td>Southeast Asia Digital Library <a href="http://sea.lib.niu.edu/reslanguage.html">http://sea.lib.niu.edu/reslanguage.html</a></td>
</tr>
</tbody>
</table>

**Table 4: The Ranking by the Performance Evaluations**

### VII. Total Results of Seven Evaluations

The Content, Usability, and Performance Evaluation (CUPE) Criteria were suggested to find which digital libraries are well-designed digital libraries among existing digital libraries in fifteen subject areas. The well-designed digital libraries in various subject areas will consist of the proposed International Open Public Digital Library (IOPDL). With the CUPE criteria, three main evaluations are performed with seven sub evaluation criteria. First, to find candidates of well-designed digital libraries, Professor McDonough and I evaluated content’s quality of existing digital libraries in fifteen subject area. We investigated: whether existing digital libraries’ content is accuracy in wide and deep scopes; whether they have
authority; and whether they give subjective Satisfaction. Through the content evaluation, sixty two digital libraries were recommended as candidates of well-designed digital libraries. Next, sixty two candidate digital libraries were evaluated again with four Usability evaluations: Accessibility evaluation with seven accessibility evaluation tools, and Interface Usability evaluations with Convenience/Ease of Use, Interfaces’ Consistency, and Visible Design and Aesthetic Appeal Evaluation Criteria. The interface usability evaluations were done by heuristic evaluation method. Lastly, to evaluate performance of sixty two candidate digital libraries, two computer programs were executed to measure the Link and Search Response Times, and Relevance of obtained results. In this chapter, I will explain how to calculate the total average for each candidate with scores of evaluations. Also, I will discuss the final results with total averages, and analyze seven evaluations with CUPE criteria for sixty two candidate digital libraries.

1. Methodology to Combine Seven Evaluations

Each evaluation is scored with 5-point scale based on each scoring method. The scores of all evaluations except content evaluation should be calculated as a center meaningful value for each candidate. The method should be able to distinguish well-designed digital libraries (WDDLs). There are several methods to summarize data, such as mean, median, mode (Weisberg, 1992). The mode is the value that appears most often in a set of data (CIMT). But, in the paper, the mode is not chosen, because it cannot give reliable results. For example, if six evaluated scores except content evaluation vary like (0, 1, 2, 3, 4, 5), the mode is unclear as #NIA in excel that I use. Also, if a set of data is (2.142857143, 4, 5, 4, 2, 2), the mode value is 4 and 2. In excel, it chooses 4 or 2 ambiguously. Because of the ambiguousness, the mode is not chosen in the paper. Also, median which finds the middle value is not chosen, because the median values of many digital libraries are not sensitive enough to differentiate data of 5 point scale. For example, 15 digital libraries out of 30 digital libraries have the same median value, 5, with six evaluated values. It makes impossible distinguish well-designed digital libraries out of sixty two candidate libraries. Generally the most effective way of summarizing the center of data is to average the values on the variable. To get a unique value for each digital library, thus, the mean is chosen. It fits for our purpose to find WDDLs. The weakness of mean is sensitive to extreme values. But, there is no extreme value in 5 point scale which may affect results.

To put together all scored values (scores) of six evaluations with the CUPE criteria, an average is calculated for each digital library. I call it as ‘final average.’ In here, the content evaluation value is excluded, because through the evaluation, the candidate, sixty two digital libraries, are recommended. Thus, six scored values of six evaluations are calculated into a final average. There are several ways to calculate the final average. The first way is to give weight to one or two specific evaluations. The second way does not give any weight to accessibility, three interface usability evaluations, or two performance evaluations. It is shown in Equation 7. The results of two ways show some different consequences to find well-designed digital libraries. Through the result of the second method, I found this way gives equal importance to each evaluation, and shows more detail weakness of the candidate libraries. And it makes me realize how much each evaluation is important to assess synthetically digital libraries. Thus, I use the second method to find well-designed digital libraries. It gives weight equally for all six evaluations: Accessibility, Convenience, Consistency, Visual design, Response time, and Relevance evaluations.

\[
\text{Final Average of each digital library} = \frac{\sum_{i=1}^{7} A_i + \sum_{i=1}^{3} U_i + \sum_{i=1}^{2} P_i}{n}
\]

Where, \(A_i\) = an evaluated value of seven values by seven Accessibility evaluation tools, 
\(U_i\) = an evaluated value by Convenience, Consistency or Visual Design Evaluation, 
\(P_i\) = an evaluated value by Response time or Relevance Evaluation, 
\(n\) = number of evaluations, that is, 6.

Equation 7: Final Average for Each Digital Library
With the final averages of the candidate digital libraries with the second way, a total average is calculated by all final averages of sixty two digital libraries, as equation 8 shows. I call it as ‘total average.’

\[
\text{Total Average of all candidate digital libraries} = \frac{\sum_{i=1}^{62} FA_i}{n}
\]

Where, \( FA_i \) = a Final Average of a candidate digital library,
\( n = \) total number of the candidate, that is, 62.

**Equation 8: Total Average of All Candidate Digital Libraries**

Then, the digital libraries that have higher average values than the total average of all sixty two candidate digital libraries are defined as well-designed digital libraries, as the equation 9 shows.

\[
\text{Final Average of a Well − Designed Digital Library} \geq \text{Total Average of all candidate DLs}
\]

**Equation 9: Final Averages of Well-Designed Digital Libraries**

For example, in the case of International Children’s Digital library, scores by evaluations are shown in the Table 5. To calculate the final average, six values in gray boxes are used.

<table>
<thead>
<tr>
<th>Seven Accessibility Evaluations</th>
<th>Average of Accessibility</th>
<th>Convenience Evaluation</th>
<th>Consistency Evaluation</th>
<th>Visibility Design</th>
<th>Response Time</th>
<th>Relevance</th>
<th>Final average</th>
</tr>
</thead>
<tbody>
<tr>
<td>F A E Say C W 3 V AE X Fujitsu</td>
<td>2.357</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>3.8928</td>
</tr>
</tbody>
</table>

Table 5: International Children’s Digital library scores

The final average = \((2.357+5+5+5+5+1)/6\) = 3.8928,
The total average = 3.2003, and
Thus, we can tell that International Children’s Digital library is a well-designed digital library.

### 2. Total Combined Results of Seven Evaluations

With the second way, the final averages of sixty two candidate digital libraries are shown in Appendix A, along with scores of six evaluations and with seven scores of accessibility evaluations. Table 6 shows sample results of sixty two digital libraries. The light gray and gray shading in the final average mean that the final average is higher than the total average. That is, the library is the well-designed digital library (WDDLS). The light gray library was/is a WDDL in the first and second average methods. The gray library means the library become a new WDDL unlike the first result, not a WDDL.

As a result, the total average of all candidate digital libraries is 3.2003 in 5-point scale. It is higher than a half of 5 points. It may mean generally that the candidate digital libraries may provide enough quality usability and performance. They are recommended by content evaluation in their subject domains. Ultimately, thirty four digital libraries out of sixty two digital libraries show higher final averages than the total average, 3.2003. That is, about half of the candidate digital libraries are turned out as well-designed digital libraries in their subject domains.

NASA’s Visible Earth (4.452381) and Census Atlas of the United States (4.263889) in Geography subject area show the highest averages by seven evaluations with the CUPE criteria. Third, U.S. Department of
Health & Human Services shows 4.261905 in medicine subject domain. Fourth, GPO Access in political science and law shows 4.142857. Fifth, National Science Foundation in science shows 4.138889. Table 6 and Appendix A show more detail results based on scores.

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>The Recommended DL by Correct Evaluation</th>
<th>Accessibility Evaluations by seven tools</th>
<th>Interface Evaluations</th>
<th>Usability Evaluations</th>
<th>Performance Evaluations</th>
<th>Final Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PAR</td>
<td>Cytheria Sa cpeoff</td>
<td>WOC</td>
<td>MASS</td>
<td>MAUR</td>
</tr>
<tr>
<td>Education</td>
<td>Exploration Digital Library</td>
<td>3.5</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>History/Thesaurus</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>SMETE digital story</td>
<td>1.5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>National Agricultural Research Digital Library</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Agriculture</td>
<td>Core Historical Library of Agricultural in China university</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>Western Digital Library</td>
<td>1.5</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Table 6: Sample Results of All Evaluations with the CUPE criteria

3. Total Analyses of Seven Evaluations

The Analysis Based on Percentages of the Final Averages

Figure 19 shows percentages of the final average of the candidate digital library. As the results, 53.968% of sixty two digital libraries gain over 3.2003 final averages. And those digital libraries are turned out as well-designed digital libraries, because their averages are higher than total average, 3.2003. 15.87% of them, 10 digital libraries, gain the final average between 3.0 and 3.2003, which couldn't be well-designed digital libraries.

The Percentages of the Final Averages in 5-point scale

Figure 19: The Percentages of the final average of all candidate digital libraries
The Analysis Based on Averages of Subject Areas

Especially, in Geography subject domain, all three candidates turn out as WDDLs: NASA's Visible Earth, Census Atlas of the United States, and David Rumsey Map Collection. Moreover, the candidate digital libraries in History of the America subject area are seven, and six digital libraries are turned out as well-designed digital libraries except Digital Past which has problems in accessing the website. According to the results based on averages of subject areas, the average of Geography subject area is the highest (4.1316). Then, the averages of Medicine (3.5571), Military Science (3.496), Political Science and Law (3.4524), Science (3.4454), Agriculture (3.3849), and Art (3.3577) are high. Ten subject areas, 66.67% of fifteen subject areas, have higher averages than the total average. On top of that, all averages of fifteen subject areas are higher than 2 points out of five points. The social science subject area is the lowest, 2.2341 average. The results show the qualities of sixty-two candidate digital libraries in fifteen subject areas are good in Content quality, Usability quality, and Performance quality overall.

![Figure 20: Comparing the Average based on Subject Domains](image)

Synthetically, I can conclude that the U.S. has strongly specialized digital libraries in Geography subject area (4.1316), Medicine (3.5571), Military Science (3.496), Political Science and Law (3.4524), and Science (3.4454). One more thing I would like to point out is that the prototype evaluations are mainly done with the U.S. digital libraries except few other country libraries such as The National Archives, Education Resources UK, British Library Online Gallery, International children’s Digital Library, and Chinese Philosophical Etext archive. However, for the International Open Public Digital Library (IOPDL), we should include and evaluate more national digital libraries of many countries.

4. Well-Designed Digital Libraries

The turned out WDDLs in fifteen subject areas by seven evaluations are listed as below. Unfortunately, in Music and books on music, and Philosophy, psychology and religion subject areas, there are no WDDLs. Thus, in 13 out of 15 subject areas, 34 out of 62 digital libraries turn out as well-designed digital libraries.

Based on Subjects Areas, the Well-Designed Digital Libraries

- Geography (4.1316)
Based on Scores, the Well-Designed Digital Libraries

<table>
<thead>
<tr>
<th>Subject Areas</th>
<th>Well-designed Digital Libraries</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Geography</td>
<td>Census Atlas of the United States (4.263889)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>4.</td>
<td>Political science, and law</td>
</tr>
<tr>
<td>5.</td>
<td>Science</td>
</tr>
<tr>
<td>8.</td>
<td>Agriculture</td>
</tr>
<tr>
<td>9.</td>
<td>Language and literature</td>
</tr>
<tr>
<td>10.</td>
<td>Language and literature</td>
</tr>
<tr>
<td>12.</td>
<td>Medicine</td>
</tr>
<tr>
<td></td>
<td>World History and history of Europe (Asia, Africa, Australia, New Zeal, and ETC)</td>
</tr>
<tr>
<td>16.</td>
<td>Arts</td>
</tr>
<tr>
<td>17.</td>
<td>Geography</td>
</tr>
<tr>
<td>18.</td>
<td>Political Science and Law</td>
</tr>
<tr>
<td>20.</td>
<td>Social Sciences</td>
</tr>
<tr>
<td>21.</td>
<td>Technology</td>
</tr>
<tr>
<td>22.</td>
<td>Military science</td>
</tr>
<tr>
<td>23.</td>
<td>Medicine</td>
</tr>
<tr>
<td>27.</td>
<td>History of the America</td>
</tr>
<tr>
<td>28.</td>
<td>Political Science and Law</td>
</tr>
<tr>
<td>29.</td>
<td>Agriculture</td>
</tr>
</tbody>
</table>
VIII. Conclusion and Discussion

The International Open Public Digital Library (IOPDL) is proposed as a global national digital library for the public and even the disabled to access any information in various subject areas. The proposed IOPDL will consist of well-designed digital libraries in fifteen subject areas by cooperation with them. The way will reduce huge amount of budget saving huge effort and time to establish it. The IOPDL can be used as an education community that provides life-long learning opportunities for the disabled and for all age with cutting-edge technologies.

To establish the proposed International Open Public Digital Library (IOPDL), it is necessary to find which digital libraries are well-designed digital libraries. To find out well-designed digital libraries, Content, Usability, and Performance Evaluation (CUPE) criteria is suggested to evaluate multiple performances of existing digital libraries. To evaluate existing digital libraries and to evaluate performance of the suggested CUPE, several evaluation methods are selected based on evaluation criteria. By seven evaluations with the CUPE criteria, 34 well-designed digital libraries are selected among 62 candidate digital libraries in fifteen subject domains. First, through the content quality evaluation, sixty two digital libraries are recommended as candidates of well-designed digital libraries in fifteen subject areas. The Content quality evaluation was done by faculty and me in Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. We investigated Contents’ accuracy, authority, and scope of collections of existing digital libraries based on their subject domains. Second, to evaluate Accessibility of sixty two candidate digital libraries, seven Accessibility Evaluation Tools are chosen based on various Accessibility standards. The method for evaluating accessibility gave enough credible results. Further, to evaluate interfaces’ Usability, the heuristic evaluation method was selected. I evaluated Interfaces’ Usability with Convenience, Interfaces’ Consistency, and Visible Design and Aesthetic Appeal Evaluation Criteria. Each criterion has their unique check lists that enable enough assessments in various points of views. Through the evaluations, I assess general interfaces’ usability of the candidate digital libraries. Third, to evaluate Performance, two computer programs are designed and executed. Through the programs, accurate results of response times and relevance rates can be easily achieved in a short time rather than using other evaluation methods. Fourth, the final average of each digital library is compared with the total average of sixty two digital libraries to decide which digital library can be well-designed digital libraries. As a result, thirty four digital libraries show higher final averages than the total average. They turn out as well-designed digital libraries with enough evidences in Content, Usability, and Performance.

Consequently, seven evaluations with the CUPE criteria are very effective and efficiency to find well-designed digital libraries, and to evaluate multiple performances of many digital libraries together. Moreover, the suggested CUPE criteria are suitable to evaluate Contents’ quality of collection, Interfaces’ Usability, and Performance for sixty two candidate digital libraries. The selected thirty four well-designed digital libraries in fifteen subject areas may consist of the International Open Public Digital Library. It will be a cornerstone to establish the proposed IOPDL. However, I should say that there are many digital libraries that are not evaluated in the prototype. With other criteria, the digital libraries that are not WDDLs in the prototype can be Well-Designed digital libraries. For the future content quality evaluation, it is strongly recommended that many experts of subject domains be involved in evaluating Content quality of existing digital libraries. It will increase possibility to find more many high quality content digital libraries in many subject areas. For the future Interface usability evaluation, it is desired that more users are involved in the evaluation. It might give better objective majority assessments for Interface Usability evaluation. But, it is also true that users cannot evaluate many digital libraries with the same
criteria without direct rewards. In the performance evaluation, synonyms and combination words as queries are not implemented to evaluate response time and relevance of query results. If synonyms and combination words are considered as queries, we can increase performance of search engines.

For the next step, therefore, it is necessary for thirty four well-designed digital libraries to cooperate solving interoperability problems. Further researches are required to achieve practically interoperability among well-designed digital libraries. For it, another prototype is planned to build interoperability among three universities libraries with a Common Terminology. A Common Terminology can be designed with several metadata schemas (e.g., MARC, MODS, and DC & QDC). If we can achieve practically interoperability among universities’ digital libraries with the Common Terminology, the achieved interoperability method can be expanded and be generalized to achieve interoperability among well-designed digital libraries. Ultimately, through the generalized interoperability method, the proposed International Open Public Digital Library can be established with cooperation with well-designed digital libraries. Further, we can expand the scope of the IOPDL involving many national well-designed digital libraries all over the world, to realize the dream that anyone can access and find any information of the world with their privilege rights.

Reference


UIUC. (n.d.). *Functional accessibility evaluator 1.0.3.* Retrieved from University of Illinois at Urbana-Champaign: http://fae.cita.uiuc.edu/


WAEX. (n.d.). *Web Accessibility Evaluator in a single XSLT file (WAEX).* Retrieved from http://www.it.uc3m.es/vle/waex.html


## Appendix A - Results of All Seven Evaluations with the CUPE criteria

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>The Recommended DLs by Content Evolution</th>
<th>Accessibility Evaluation by seven tools</th>
<th>Interface Usability Evaluation</th>
<th>Interface Usability Evaluation</th>
<th>Interface Usability Evaluation</th>
<th>Interface Usability Evaluation</th>
<th>Performance Evaluation</th>
<th>Final Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>The National Library of Education</td>
<td>3.5</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>Exploration Digital Library</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="http://explorationdigital.library.org/">http://explorationdigital.library.org/</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History</td>
<td>History Maine</td>
<td>4.1</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>4.5</td>
<td>2</td>
<td>2.71428714</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.historymaine.com/">http://www.historymaine.com/</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>J-MDL Digital Library</td>
<td>1.5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2.5</td>
<td>0</td>
<td>2.071428571</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.jmdl.or">http://www.jmdl.or</a>                g/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculture</td>
<td>North Carolina Agricultural Library</td>
<td>2.2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2.5</td>
<td>0</td>
<td>1.357142857</td>
</tr>
<tr>
<td></td>
<td><a href="http://library.cnc.ncsu.edu/">http://library.cnc.ncsu.edu/</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>University of Nebraska</td>
<td>1.5</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>2.5</td>
<td>2</td>
<td>2.571428571</td>
</tr>
<tr>
<td></td>
<td><a href="http://library.unl.edu/">http://library.unl.edu/</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NASA's Visible Earth</td>
<td>3.1</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2.071428571</td>
</tr>
<tr>
<td></td>
<td><a href="http://visibleearth.nasa.gov/">http://visibleearth.nasa.gov/</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geospatial</td>
<td>David Rumsey Map Collection</td>
<td>3.5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>4.5</td>
<td>2</td>
<td>3.714285714</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.davidsmap.com/">http://www.davidsmap.com/</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Census Atlas of the United States</td>
<td>3.5</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2.071428571</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.census.gov/">http://www.census.gov/</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NYPL Digital Gallery</td>
<td>4.5</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>4.5</td>
<td>3 fail</td>
<td>3.571428571</td>
</tr>
<tr>
<td></td>
<td><a href="http://digitalgallery.nypl.org/">http://digitalgallery.nypl.org/</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arts</td>
<td>Museum of Modern Art</td>
<td>3.1</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>2.214285714</td>
</tr>
<tr>
<td></td>
<td>Art History</td>
<td>3.5</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>2.714285714</td>
</tr>
<tr>
<td></td>
<td>Resources on the Web</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="http://artscience.bsc.edu/">http://artscience.bsc.edu/</a>            KTH/Art.4168.geo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>William Blake Archive</td>
<td>3.2</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>2.5</td>
<td>3.257142857</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.blakearchive.org/">http://www.blakearchive.org/</a>          s/blakearchive/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

49
<table>
<thead>
<tr>
<th>Subject Area</th>
<th>The Recommended DLs by Content Evaluation</th>
<th>Accessibility Evaluations by seven tasks</th>
<th>Interface Usability Evaluations</th>
<th>Performance Evaluations</th>
<th>Final Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>National Science Digital Library <a href="http://dsdl.org/">http://dsdl.org/</a></td>
<td>4.5</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>US. Department of Energy <a href="http://www.energy.gov/energy/index.html">http://www.energy.gov/energy/index.html</a></td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>National Science Foundation <a href="http://www.nsf.gov/">http://www.nsf.gov/</a></td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2.5</td>
</tr>
<tr>
<td>Science Photo Library <a href="http://www.sciencephotolibrary.com/index.html">http://www.sciencephotolibrary.com/index.html</a></td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2.5</td>
</tr>
<tr>
<td>Analytical sciences Digital Library (ASEL) <a href="http://www.anlib.net/find-a.html">http://www.anlib.net/find-a.html</a></td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>2.5</td>
</tr>
<tr>
<td>Public Library of Science <a href="http://www.plouglurgy.org/education.action">http://www.plouglurgy.org/education.action</a></td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DL@II 4</td>
<td>2.5</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>3.5</td>
</tr>
<tr>
<td>DarmstadtMIT <a href="http://darmstadt.mit.edu/">http://darmstadt.mit.edu/</a></td>
<td>3.3</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>4.5</td>
</tr>
<tr>
<td>JISC Digital Image Collection <a href="http://images.jisc.ac.uk/govindex.html">http://images.jisc.ac.uk/govindex.html</a></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2.5</td>
</tr>
<tr>
<td>IEEE computer society <a href="http://www.computer.org/">http://www.computer.org/</a></td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2.5</td>
</tr>
<tr>
<td>History and history of Europe (Ass)</td>
<td>3.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Digital Library Online Catalog <a href="http://www.bl.uk/onlinegallery/index.html">http://www.bl.uk/onlinegallery/index.html</a></td>
<td>3.5</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>EuroDocs <a href="http://eurodocs.lib.byu.edu/ezproxy.byu.edu/libdocs.php?user=">http://eurodocs.lib.byu.edu/ezproxy.byu.edu/libdocs.php?user=</a></td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Africa Focus Sights</td>
<td>3.5</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
Results of All Seven Evaluations with the CUPE criteria