
This version is available at https://strathprints.strath.ac.uk/2622/

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (https://strathprints.strath.ac.uk/) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to the Strathprints administrator: strathprints@strath.ac.uk

http://eprints.cdlr.strath.ac.uk/2622/

This is an author-produced version of a paper from the International Conference on Digital Libraries, 27-27 Feb 2004.

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in Strathprints to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profitmaking activities or any commercial gain. You may freely distribute the url (http://eprints.cdlr.strath.ac.uk) of the Strathprints website.

Any correspondence concerning this service should be sent to The Strathprints Administrator: eprints@cis.strath.ac.uk
Access to Information in Digital Libraries: Users and Digital Divide

G.G. Chowdhury

Graduate School of Informatics, Department of Computer and Information Sciences
University of Strathclyde, Glasgow, UK
e-mail: gobinda.chowdhury@cis.strath.ac.uk

Keywords
Digital libraries, Digital divide, Information access, Information users, Information society

Abstract
Recognising the importance of information and knowledge in all spheres of human life, the recently held World Summit on Information Society came up with a plan of action for building a global information society. The goal of the world information society initiatives is the same as that of digital library research and development – to make information and knowledge accessible to everyone in the world. Digital libraries have progressed very rapidly over the past ten or so years. This paper addresses the two most important aspects of the information society – information users and digital divide. Findings of some large-scale studies on human information behaviour on the web and digital libraries have been discussed. The major findings of a study on access to electronic resources by university students are the presented. Proposed that a one-stop window approach with a task-based information organisation and access system may be the way forward.

Introduction

“We, the representatives of the peoples of the world, assembled in Geneva from 10-12 December 2003 for the first phase of the World Summit on the Information Society, declare our common desire and commitment to build a people-centred, inclusive and development-oriented Information Society, where everyone can create, access, utilize and share information and knowledge …..” (Draft Declaration of Principles, World Summit of the Information Society. WSIS/PC-3/DT/6(Rev 3)-E, 9 Dec. 2003)

This is the vision of the Information Society as described in the ‘Draft Declaration of Principles’ of the World Information Summit, Geneva, December 2003. The document further states: ‘We recognize that education, knowledge, information and communication are at the core of human progress, endeavour and well-being.’ (Draft Declaration of Principles 2003) It goes on describing the various challenges of building an information society and proposes a number of approaches. At one place it states that ‘all stakeholders should work together to: improve access to information and communication infrastructure and technologies as well as to information and knowledge ….’ Aren’t we all – people engaged in digital library research and development – working towards this goal – to make information and knowledge accessible to all? Digital library research and development activities have a key role to play in building a global information society.

Interests in digital libraries, both scholarly and professional, grew very rapidly in the 1990s. As a result of digital library research in the US, UK, Europe and in other parts of the world, a
number of digital libraries have been established during the past ten or so years. These digital libraries differ in size, collection, features and services. Have the digital libraries reached the stage of meeting the vision of the global information society? If not, what needs to be done? This paper addresses these issues, and focuses mainly on two of the most important issues of the Information Society, viz. information users and digital divide.

**Research Problems: Information Users and Digital Divide**

An information user in today’s digital world has to find the required information from various distributed information sources. Reviews of digital libraries show the diversified collection of information resources available to users ranging from full-texts of journal and conference papers to CD-ROM databases, theses and dissertations, e-journals, e-books, examination papers, images and photographs, maps, audio, video and multimedia resources, manuscripts, and so on. (Chowdhury and Chowdhury 2000; Meyyappan, Chowdhury and Foo 2000)

Apart from those that are currently managed, and provided access to, by libraries there are many other information resources that are needed by users in an academic environment. These resources include course calendars, university statutes, various course information, thesis and dissertation guidelines, style guides, laboratory facilities, availability of software, hardware, equipment, course materials, and so on. Currently digital libraries do not manage or provide access to these diverse but yet extremely useful collections of information.

Access to electronic information is significantly influenced by digital divide – the perceived disadvantages of those who are either unable, or do not choose, to use the appropriate ICT to perform their activities, decision making, learning and pleasure (Cullen 2001, Digital Divide network). Digital divide is a major area of research around the world, and cover many different disciplines, individuals and institutions including national and international bodies. Details of these research activities can be found in a number of sources, for example general works on digital divide including statistics, etc., can be found on the Digital Divide Network, and issues concerning digital divide and libraries can be found in a number of works listed in the bibliography compiled by Ryan (n.d.).

This paper addresses some major issues and problems related to the human information behaviour in a digital environment, and proposes a user-centred design approach for the future digital libraries.

**Access to Information in Digital Libraries**

Digital library resources can be accessed in a variety of ways, for example through the:

- Institution’s library webpages
- Webpages of specific digital libraries such as NDLTD, NCSTRL, NZDL, ACM digital library, PubMed
- Subject Gateways, such as SOSIG, Biz/ed, OMNI
- Search service providers such as Dialog, Ovid online, Proquest
- Web search tools: search engines including meta search engines and speciality search engines like Google, AskJeeves, Kartoo, Vivisimo, etc., and directories like Yahoo.

General and specific information retrieval features of digital libraries have been discussed in a number of documents (see for example, Chowdhury and Chowdhury 2000, Chowdhury and Chowdhury 2003, and Chowdhury 2004). Access to information in digital libraries is influenced
by a number of factors. Several researchers have studied human information behaviour on the
web, and the findings of such studies are quite relevant to digital libraries.

**Human information behaviour on the web: some recent findings**

Information retrieval services on the web deal with text as well as multimedia information
resources that are linked with other documents, and there is no target user community as such.
Researchers have noted that the retrieval characteristics of web information services are designed
for the general audiences, and they are different from the features of the traditional IR systems
studied the transaction logs of AltaVista for six weeks with over a billion search requests, and
noted that an average query size was 2.35 terms. They also noted that most search sessions were
short: in 77% of the sessions only one search request was made and in 63.7% of the sessions only
one query was entered and one result screen viewed. After analysing about 16 million queries from the
German search engine Fireball, Hoelscher (1998) noted that only 2.6% queries used Boolean operators and
over 54% queries had only one search term. They noted that there were no statistical differences in the retrieval effectiveness of
eight search engines viz. Yahoo, Excite, Magellan, HotBot, AltaVista, Lycos, Infoseek and Open Text. They
noted that majority of the web searchers are not comfortable with Boolean query formulation, and they do not frequently browse the results. Tenopir
(2003) reviews more than 200 recent research publications – published between 1995 and 2003 –
that focus on the use of electronic library resources. Amongst others, she noted that college and
school students use the Internet more than the library for research, and often the quality
judgements that they make about the retrieved materials are not adequate.

Cooper (2001) analysed the usage of the Melvyl web catalogue at the University of California
at Berkeley. During the 479 day study period users conducted about 2.5 million search sessions,
during which about 7.4 million search statements were executed. This study revealed that:

- The length of time the searchers spent at a session grew from 6 to 10 minutes during the 16
  months study period.
- During the study period about 7.4 million database selections were made, out of which
catalogue database accounted for 32% (about 2.4 million uses), Medlars for 22%, magazine
database 10%, periodicals database 8%, Inspec 4% and Current Contents 4%.
- The time users spent on each database also varied; for instance about 3 minutes per search in
  the Medlars database and 2.2 minutes for the catalogue database.
- Users displayed about 4–5 citations for each search they performed.
- The length of time users spent displaying results was 30–40 seconds.

Table 1 summarises human information behaviour across different search systems. In a more recent
study with the AltaVista search engine Ford, Miller and Moss (2003) noted that:

- Boolean only searches were associated with poor retrieval.
- Boolean only searches performed worse than best match and combined searches.
- Combined searches performed worse than best match alone searches.
- There was no evidence that these results were related to errors, choice of term or any interaction between
  them, in Boolean as opposed to best match or combined searches.
Table 1: User search patterns in different online search environments

<table>
<thead>
<tr>
<th>Issues</th>
<th>Web search systems</th>
<th>Traditional IR systems</th>
<th>OPAC search systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Queries per user per session</td>
<td>1–2</td>
<td>7–16</td>
<td>2–5</td>
</tr>
<tr>
<td>Terms per query</td>
<td>2</td>
<td>6–9</td>
<td>1–2</td>
</tr>
<tr>
<td>Relevant documents viewed per session</td>
<td>10 or fewer</td>
<td>approx. 10</td>
<td>fewer than 50</td>
</tr>
<tr>
<td>Queries containing advanced options</td>
<td>9%</td>
<td>9%</td>
<td>8%</td>
</tr>
<tr>
<td>Queries containing Boolean operators</td>
<td>8%</td>
<td>37%</td>
<td>1%</td>
</tr>
<tr>
<td>Queries improperly formatted</td>
<td>10%</td>
<td>17%</td>
<td>7–9%</td>
</tr>
</tbody>
</table>

In summary, the evaluation studies reported above reveal that:

- Users find it difficult to formulate complex queries.
- In general users spend very little time on searching a given web search tool or database.
- In most cases users formulate very short and simple queries with one or two search terms and very few search operators.
- Users spend very little time in looking at and deciding the usefulness or relevance of retrieved items.
- Very few queries contain advanced search features.

Digital Divide: Myth and Reality

Digital divide is often measured by the degree of access to ICT and the Internet. Indeed the figures for accessibility vary between the reach and poor countries, and even among the rich and the poor within a developed country. For example, in 2001, in the US total home Internet access was 55% for whites compared to 31% for African-Americans and 32% for Hispanics (Warschauer 2003). While improved access to ICT and the Internet are important for all sorts of reasons, one should not think that computers, networks or the Internet can alone solve all the problems. In fact, mere access to ICT and the Internet does not mean that these will be used appropriately to access information and knowledge for solving problems, or for education, learning, research, and so on. One should not forget that these are mere tools and they need to be used appropriately to get access to the required information, and digital library researchers and managers have a great role to play to ensure that this happens. Warschauer (2003) comments that ‘Technology does not exist as an external variable to be injected from the outside to bring about certain results.’ Furthermore access to the appropriate ICT and the Internet may provide access to the information resources, but finding the most appropriate information relevant to solve a problem still remains. Designers of digital libraries can make the appropriate use of the technology to facilitate retrieval of the most relevant information with the minimum user efforts. This is one of the major challenges for today’s digital library researchers.

Problems of Using Electronic Libraries: Preliminary Results of a Research Project

Academic users are probably the largest and the most frequent users of digital library services. Most university libraries today provide access to a range of printed as well as digital
information resources. University students, staff and researchers can now access a variety of digital resources ranging from e-journals and e-books to electronic databases, local and remote digital libraries and the world wide web. However, it is important to know how easily a user can get access to all the information resources relevant to his/her query. With a view to assessing the ease (or difficulties) of getting access to electronic information through the university library webpage, an experiment was conducted at Strathclyde with 55 users. Detailed findings of this study will be reported elsewhere; a brief description of the study and some major findings are reported in the following sections.

The Experiment

Fiftyfive students from the MSc. Information and Library Studies class were chosen for this study. Each student was asked to choose a topic of his/her research interest and was asked to use the Strathclyde University library web to find information resources relevant to the topic. There were altogether 55 topics. The following examples show the kind of research topics chosen by the users:

- Marketing of public library services
- Intellectual property laws in Britain
- Harry Potter book series: its impact on library readership and protests from religious groups
- e-government: implementing electronic elections and democracy
- Digital libraries and their use
- Managing digital libraries
- Influence of the horror genre on modern cinema
- Privacy and ethics in an electronic age

In order to make the searches realistic, no restrictions were imposed on the users’ search time, and the choice of the type of information resources and/or services available through the library webpages. Instead they were asked to find all the possible resources ranging from book materials (which could be found through the library OPAC) to journal and conference papers which could be found through a number of e-journals and electronic databases, theses and dissertations that could be searched through the university thesis collection, and all other types of resources that could be accessed through the websites or search tools pointed to by the library webpage. Users were asked to stop when they considered that a reasonable number of resources of all the different types were found. While searching for information using a specific system e.g., e-journals or electronic databases, users could stop when they had thought that enough time was spent on it, even if the desired results were not obtained. Thus, the objective was not to compare the performance of the users or of one or more specific systems, but to measure the overall usability of the digital (strictly speaking hybrid) libraries of today. The users were asked to note down their experience, especially the level of difficulties faced while conducting the searches, and were asked to comment on each search system and the overall digital library.

Findings

A qualitative research method was followed for this study. The following are some of the most important findings, obtained as the users’ comments, of this study.

1. An online search calls for some basic ICT and information literacy skills without which a search may never be fruitful.
2. Initial searches may not produce any good hits, yet there may be a wealth of information resources that are relevant to a search topic.
3. A particular search may produce too many or too few hits; users should be prepared to spend time and try out the various alternative search options.
4. Different systems – online search service providers, e-journals and aggregators, web search tools, and digital libraries – have different search interfaces and often use different search syntax; this makes the learning curve very steep.
5. Familiarity with the search topic and the structure of the digital library as well as the organisation and content of the various systems – e-journals, online databases, digital libraries, etc., helps the user produce better search results at a relatively less time.
6. There may be a number of e-journals and databases covering a given subject/topic. It is often difficult for the user to decide which journal or database to select, and yet this is the first step and it determines the choice of the appropriate search service provider (for electronic databases) or aggregator (for e-journals).
7. Searches through online databases, e-journals, etc., often produce only abstracts, and the user has to go through another process to get access to the full text of the retrieved items.
8. Some search options, e.g., ‘keyword’, ‘subject’, ‘title keyword’, and ‘subject keyword’, are often confusing; users may not know the differences.
9. Off-campus searching is often very time consuming and frustrating, because due to the low bandwidth connection from the home computer, a search and download may take a long time, and in the meantime the connection may be broken. On many occasions, the user is prompted to enter the password to re-establish the connection. All these add to the frustration of the user.
10. Sometimes a search produces too many hits and the user has to spend a lot of time (for example for full text journal articles) to decide whether or not an item is relevant.
11. A lot of useful information is available on the departmental and faculty intranet sites in the form of lecture notes and handouts, useful reading materials and pre-prints, suggested web sites and digital resources, etc., which are not accessible through the library web pages.
12. Organization of information on the main library page has an impact on the user’s selection of a specific digital library service; sometimes the specific services are hidden under several layers, and the user may have to explore the library webpage to discover the various resources and services available.
13. Given the multiple variety of information resources, it is often difficult to decide which one to use, and to decide the most relevant items retrieved through all the various systems is a challenging task.
14. Once the user chooses to search on a specific service, e.g. e-journals, or electronic databases, it is difficult to come back to the library web page and start another search; results from one particular search have to be saved, the user has to come back to the library webpage, and choose another system to search; often the user has to go through the whole process again, e.g., logging-in with the user id and password, etc.
15. Selection of appropriate search terms is a big challenge, especially for a complex or an unknown topic, and yet this is the first step in any search.

Possible Solutions
The above findings highlight some of the most common problems faced by today’s digital library users. Several design and user issues are involved here, for example,

1. *Point 1:* This has direct implications on digital and information literacy training; indeed digital and information literacy training are not adequate in most universities, and yet the complexities of the digital information world call for regular and rigorous training.
2. *Points 2, 3 and 4:* These mean that users should be prepared to spend more time to search for information, and this is in contrast with the typical web search behaviour of users where, as discussed earlier in this paper, a typical search takes only 2-3 minutes. This has to be stressed in any user education or information literacy training; this is also an important design consideration.

3. *Point 4:* A one-stop window approach where the user can see and use only one search interface to search information from a variety of systems may resolve this problem.

4. *Point 4:* A task-based information access system may resolve this problem; alternatively there should be enough online help to guide the user.

5. *Point 5, 6, 7 and 8:* These have implications on information literacy, user education and online help facilities.

6. *Point 9:* This is directly related to the digital divide. Most recent statistics show only whether a household has an Internet connection or not. However, in order to make it worthwhile for access to and use of digital libraries, it is important to have a high bandwidth connection.

7. *Point 10:* This is an important point since from web search studies (reported earlier in this paper) it is noted that the user does not spend much time on looking at the search output. Automatic filtration mechanisms based on user characteristics, user tasks and user choices, may be useful.

8. *Point 11 and 12:* A task-based information access system may be designed to provide access to the library as well as the institutional intranet and internet resources.

9. *Point 12, 13 and 14:* Cross-database searching facilities with options for discarding duplicates and ranking the results (based on a user-driven criteria) may resolve the problem.

10. *Point 15:* Facilities for using search term dictionary or vocabulary control tools are essential for good digital library search interfaces.

**Need for a task-based information access through a one-stop window**

Since information relevant to a given task or a specific problem may be obtained from a variety of information systems and sources, it would certainly be helpful to design a one-stop window for access to all kinds of information systems and services. Indeed this has been pointed out by many researchers some of whose works have been discussed earlier in this paper.

Information needs of users are influenced by their nature of work, affiliation, educational background, accessibility to technology, and so on. Users face a number of problems in today’s electronic information environment. Firstly, they do not know which information source may be appropriate to accomplish a particular task or to resolve a specific problem. Secondly, even if the users are aware of the existence of a particular information source, they may not know how to locate it, and finally how to retrieve the information. Current digital information systems, such as those obtained from digital libraries or the Web, are not organised to match the various tasks that users perform. As a result, users often source for information through trial and error, and browse Web pages from one to another, or move from one information resource to another to obtain the desired information. This is unproductive and results in a waste of time and energy. Some recent research efforts have been directed to build personalised digital library environments to solve some of these problems (see for example, Adams, Judith and Sharon 1995, Barry and Barbara 1999, Suzane, et al, 2000, Zhang, 1998 and Zhao, 1998). Some studies on the user-centred digital library design have also been reported in the literature (see for example, Amato and Straccia 1999, and Theng et al 1999). However, none of these research projects has proposed a
one-stop approach to providing access to all kinds of electronic information, available through the digital/hybrid library as well as intranets, that are required by the users in an organisation.

**DWE (Digital Work Environment)**

Over the past few years, I along with my co-researchers in Singapore, have worked for the designing a task-based digital information access system called DWE (Digital Work Environment). The main objective of this design has been to provide a one-stop information access point for both internal and external information that are accessed by the different categories of users in an organisation. DWE was designed primarily for academic users, though it can be modified and used for access to digital information resources by users in any organization. DWE provides access to information resources in the in-house libraries, local and remote digital libraries, and the vast array of information resources in the academic institution’s own Intranet systems. Design details of DWE have appeared in a number of papers (Meyyappan, Chowdhury and Foo 2001a,2001b,2001c), and the evaluation aspects have been reported in a paper that is due to appear soon (Meyyappan, Foo and Chowdhury 2004). Once a user logs in, DWE can automatically display the tasks that are usually performed by the concerned user, and at this stage the user can select a specific task in order to look for the various information resources that are relevant to the chosen task. A given task may have a number of sub-tasks, and upon selecting an appropriate sub-task (leaf in the hierarchical tree), the corresponding information resources are shown in the **Display** frame. Users have different options to display these resources in the **Display** frame: by alphabetical order, by LCSH subject headings, by resource-type, date of last access or frequency of use. Different icons are used to indicate different information resources and for denoting the current status of the information resources. DWE also allows the user to conduct a search on a specific type of resource, e.g. OPAC, online databases, e-journals, etc. The user can also conduct a cross-database search in which case the user’s search expression is thrown onto a number of web or digital library sites, and results are brought back to the user.

DWE also incorporates a personal space and folder for users to store and access various information resources at a later point in time. The personal space is used to record users’ favourite or frequently used URLs. This is basically a convenient bookmark facility for users to revisit stored URLs, or to update or add new URLs over time. Links to full-text papers are stored under this option. Users can also make notes on important quotes, texts or other annotations and store them in the personal space which can subsequently be searched.

The task-based approach in DWE provides a means to achieve a user-centred paradigm, to allow users to be taken to their tasks at hand directly, and to allow them to access a filtered set of information resources to meet their needs. DWE encompasses both ‘local’ and ‘global’ characteristics in its interface to the various heterogeneous internal and external resources. It is scalable and can be implemented in similar settings with minimal reconfiguration. Wider testing and application of the DWE model can make significant contributions towards building a global information society by improving access to the global and local information.

**Conclusion**

Digital library research and development have grown quite fast over the past ten or so years, and many digital libraries have grown from small experiments to real-life digital libraries with huge collections. The focus originally was to build enabling technologies to design and build digital libraries with distributed collections of digital resources in different forms and formats, and
the story has been successful so far. Time has now come to focus on the users, evaluation and impact studies; indeed several researchers are now engaged in these studies.

Bridging the digital divide, by providing access to ICT and the Internet, is a necessary condition for achieving the goals of digital library research and development, but it is not sufficient. Building digital libraries and making them easily accessible, and ensuring their optimum use are the goals of digital library research and development activities. The closer we get to achieving these goals the more prominent will be our contributions towards building a global information society: ‘to promote the use of information and knowledge for the achievement of internationally agreed development goals’ (Draft Plan of Action 2003).

Digital library researchers should focus on developing systems that will reduce the burden of the users, allowing them to find the relevant information from all the possible systems and services with as minimum effort as possible. A one-stop window approach, with facilities for task-based organisation and access, personalization and automatic filtering based on user characteristics and tasks, may bring the solution. The best approach may be to build a layer on top of the existing array of web and digital library services that can be user-centred and can hide all the complexities of each individual system and service. Flexible, portable and scalable design layers, something like DWE, may be built and used by every institution providing digital, more specifically hybrid, library services.

Acknowledgements
The author wishes to express his sincere thanks to the students of the MSc in Information and Library Studies course, at the University of Strathclyde, Glasgow, who took part in the experiment reported in this paper.

References


