Digital libraries and reference services: present and future

Gobinda G. Chowdhury
Department of Computer and Information Sciences, University of Strathclyde, Glasgow, UK

Revised 19 December 2001

Keywords Libraries, Information services, Information technology

Abstract Reference services have taken a central place in library and information services. They are also regarded as personalised services since in most cases a personal discussion takes place between a user and a reference librarian. Based on this, the librarian points to the sources that are considered to be most appropriate to meet the specific information need(s) of the user. Since the Web and digital libraries are meant for providing direct access to information sources and services without the intervention of human intermediaries, the pertinent question that appears is whether we need reference services in digital libraries, and, if so, how best to offer such services. Current digital libraries focus more on access to, and retrieval of, digital information, and hardly lay emphasis on the service aspects. This may have been caused by the narrower definitions of digital libraries formulated by digital library researchers. This paper looks at the current state of research in personalised information services in digital libraries. It first analyses some representative definitions of digital libraries in order to establish the need for personalised services. It then provides a brief overview of the various online reference and information services currently available on the Web. The paper also briefly reviews digital library research that specifically focuses on the personalisation of digital libraries and the provision of digital reference and information services. Finally, the paper proposes some new areas of research that may be undertaken to improve the provision of personalised information services in digital libraries.

Introduction
Reference services, sometimes referred to as “reference and information services”, refer to the personal assistance provided to users in the pursuit of information (Bunge, 1999). Provision of such personalised information services has remained the central theme of the library and information profession. The importance of these services grew over time with the introduction of new technologies and services in libraries. Bunge (1999) categorised reference services into three broad groups:

1. information services that involve either finding the required information on behalf of the users, or assisting users in finding information;
2. instruction in the use of library resources and services (broadly defined as information literacy skills); and
3. user guidance, in which users are guided in selecting the most appropriate information sources and services.

Reference questions can range from the simple fact-finding type to complex questions requiring consultation, and often analysis, of one or more
information sources (Chowdhury and Chowdhury, 2001a). An important part of a reference service is the reference interview, which involves a personal discussion between a user and the reference librarian. Through the interview the reference librarian not only tries to understand the specific information need(s), but also collects information about the user, such as the user’s subject knowledge, the purpose of finding the specific information, and so on. Based on the reference interview, the reference librarian is often able to filter the retrieved information in order to pick up the most appropriate source(s) for the given user at the given point of time. While reference service is largely a reactive service, i.e. assistance or the service is provided when asked for by the users, libraries have also played a key role in providing information services in anticipation of user needs. Such services include various forms of current awareness and selective dissemination of information services. These types of service that aim to keep the users abreast of the latest developments in their areas of interest, however, have not been provided only by libraries: online search services like Dialog and institutions like the Institute of Scientific Information (ISI) have provided such services, for example Dialog Alerts, Current Contents and so on.

With the increasing availability of electronic information services, and the emergence of the Web and digital libraries, an obvious question arises: is there a need for such personalised services? To put it in the words of Arms, who is an active researcher and a strong proponent of digital libraries: “could we conceive of an automated digital library that disintermediates all the services that reference librarians now provide?” (Arms, 2000). This is a very pertinent question for the future of digital libraries as well as the library and information science profession. This is the central theme of this paper. It is obvious that the current digital libraries, and most of the digital library research projects, have focused on access to, and retrieval of, digital information rather than on the services that have been provided to the users by traditional libraries. Perhaps the reason for this is the narrower definition of digital libraries formulated by the digital library projects themselves. This paper looks at some definitions of digital libraries. However, since the definitions of digital libraries have been critically analysed by Borgman (1999, 2000), the present paper begins from there. It also looks at some definitions of digital libraries that emphasise services, implicitly or explicitly.

The Web has altered the notion of traditional reference services, and a number of Web-based “expert services” are now offered by many non-library organisations (Janes et al., 2001). This paper briefly reviews some currently available Web-based reference services. Although these reference services are not necessarily offered by libraries, and they are termed differently, for example, as “expert services”, they are used quite heavily, and they do not require a visit to the library to get answers to the questions. So, what should be the role of digital libraries? Should they stop providing reference services
altogether or should they concentrate only on certain types of reference questions? This paper looks into these issues. Current digital library research that specifically focuses on providing reference and personalised information services is also discussed. Finally, this paper points out some research areas that may be looked at for second generation digital libraries in order to change the notion of digital libraries from information access points to information service providers.

**Definition of digital libraries and personalised services**

While reviewing the definitions of digital libraries, Borgman (1999) noted that “in general, researchers (who primarily come from computer science and/or engineering backgrounds) focus on digital libraries as content collected on behalf of user communities, while librarians focus on digital libraries as institutions or services”. She further stressed that the current state of digital library research and development, especially in the USA, is influenced by the research definition of digital libraries (Borgman, 2000). The emphasis on the “enabling technologies” for access to digital resources rather than on the services traditionally provided by libraries becomes quite evident when one looks at the two special issues of the *Journal of the American Society for Information Science* (2000a, 2000b). The two issues contain excellent articles reflecting the state of digital library research (especially in the USA), but none focus on digital library services *per se*. Even the editor of the two special issues did not discuss reference and information services. He mentioned that “digital libraries have the ability to provide a previously unattainable level of service” (Chen, 2000). He emphasised access to information and suggested the various methods, tools and techniques that facilitate improved access to information in digital libraries. However, this is in sharp contrast to the views of the two editors of the special issue of *Information Processing & Management* (1999) on digital libraries. Marchionini and Fox (1999) in their editorial stated that “digital library work occurs in the context of a complex design space shaped by four dimensions: community, technology, services and content”.

The fact that digital libraries should not be regarded only as point of access to digital information was emphasised by a definition of digital libraries given by Borgman as early as in 1992 whereby she contended that a digital library (then defined as an electronic library) is “(1) a service; (2) an architecture; (3) a set of information resources, databases of text, numbers, graphics, sound, video, etc. and (4) a set of tools and capabilities to locate, retrieve and utilize the information resources available” (Borgman, 1999). The definition of a digital library that came up in the March 1994 Digital Library Workshop emphasised that a full service digital library must accomplish all the essential services of traditional libraries and also exploit the well-known advantages of digital storage, searching and communication (Gladney *et al.*, 1994). There are many other definitions of digital libraries that mention the services aspect, either
explicitly or implicitly (see for example, Oppenheim and Smithson, 1999; Rusbridge, 1998; Waters, 1998).

Pointing out the lack of emphasis on reference and information services, Sloan (1998) lamented that “human interaction in the digital library is discussed far less frequently. One would almost get the impression that the service tradition of the physical library will be unnecessary and redundant in the digital library environment”. He emphasised the point that “technology and information resources, on their own, cannot make up an effective digital library”. Emphasising the need for personalised services in digital libraries, Lombardi (2000) mentioned that, “helping clients find resources in a digitally chaotic world is the first priority”. In a Delphi study of digital libraries it was revealed that “the primary roles librarians play in digital libraries include organisation (cataloguing and indexing), selection and acquisition and acting as gateways to the provision of services involving information” (Kochtanek and Hein, 1999). This study also revealed that “the best reasons for developing a digital library include: increasing access to information, serving end users’ needs and bringing organisation to the unstructured universe of electronically available information”. Thus, this study shows the importance of “serving end users’ needs”, and the need for the librarians to act as “gateways to the provision of services involving information” in a digital library. Another study conducted by Downs and Friedman (1999) indicated that “in the absence of human intermediaries, there is a need for end-user instruction on the use of digital libraries”. Covi (1999) pointed out that “digital libraries will be used only when workers can readily integrate into social legitimated and legitimateable ways of working”. She further suggested that subject specialists, as digital library service providers, help users formulate disciplinary search strategies and provide assistance in developing new resources.

Arms (2000), while trying to answer the question “whether or not we need reference librarians in a digital library”, pointed out that automatic tools are steadily reducing the need for reference librarians in fields like medicine that required a significant amount of intermediation in information searching. The essence of Arms’ paper is that while computers can do a lot in performing most of the traditional jobs of reference librarians, their skills may be required in case of complex information searches. Arms, however, believes that there have been significant improvements in the end-user search facilities, provided by electronic database search services and Web search engines, and hence the need for human intermediaries in information searching has reduced significantly.

Perhaps the most comprehensive view of the need for and the nature of personalised services in a digital library has been delineated by Marchionini and Fox in their editorial of the special issue of Information Processing & Management (1999) on digital libraries. They maintained that:
Services reflect the functionality afforded by systems serving the community of users. Access services that facilitate search and browsing have been central to digital library research thus far, but there is great need for attention to reference and question-answering, on-demand help, and fostering of citizenship and literacy and mechanisms to simplify participatory involvement of user communities.

262

Reference and information services on the Web
A number of reference and information services are now available on the Web. Interestingly, many of these services are provided by non-library and commercial organisations. While some are free, others need payment. Detailed discussions on such services are available in a number of publications (see for example, Chowdhury and Chowdhury, 2001a, 2001b; Lankes et al., 2000; Sherman, 2000), McKiernan (2001) maintains a site that provides categorised listing of libraries that offer real-time reference services using chat software, live interactive communication tools, call centre management software, bulletin board services and other Internet technologies. Most of these services are designed for registered users of the libraries concerned.

Chowdhury and Chowdhury (2001b) categorised online reference and information services into three broad groups:

1) reference and information services from publishers, database search services, and specialised institutions;

2) reference services provided by libraries and/or experts through the Internet; and

3) reference and information services where users need to conduct a search and find information through the Web.

Chowdhury and Chowdhury (2001a) discussed several online information services that belong to the first category mentioned above. They have listed various current awareness and SDI services such as:

- the contents page service from commercial publishers, such as Elsevier’s Contents Direct Service, IDEAL Alert from Academic Press, and so on;

- information on new books available free from publishers and vendors, such as the Wiley Book Notification Service and Amazon.com;

- SDI services from online search service providers, such as Dialog (Dialog Alerts);

- Current Contents and ISI Alerting Services from ISI, and so on.

Some of these services, particularly the contents page services from publishers of journals, are free, while for others, such as Dialog Alerts or Current Contents from ISI, users need to register and pay.

Table I provides a quick overview of some online reference and information services currently available. This is not an exhaustive list, but the table shows
<table>
<thead>
<tr>
<th>Service</th>
<th>Subject</th>
<th>Payment</th>
<th>Organisation</th>
<th>Service providers</th>
<th>Question input</th>
<th>Mode of delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Askme</td>
<td>All</td>
<td>Free</td>
<td>Askme.com</td>
<td>Experts (volunteers)</td>
<td>Select a subcategory and input query through a Web-based query form</td>
<td>E-mail</td>
</tr>
<tr>
<td>AllExperts</td>
<td>All</td>
<td>Free</td>
<td>AllExperts.com</td>
<td>Experts (volunteers)</td>
<td>Select a subcategory and input query through a Web-based query form</td>
<td>E-mail</td>
</tr>
<tr>
<td>Inforocket</td>
<td>All</td>
<td>Fee-based (from $5 to $75)</td>
<td>Inforocketcom</td>
<td>Experts (volunteers)</td>
<td>Select a subcategory and input query through a Web-based query form</td>
<td>E-mail</td>
</tr>
<tr>
<td>AskAuntie</td>
<td>Law</td>
<td>Free</td>
<td>Nolo.com</td>
<td>Subject experts</td>
<td>Input query through a Web-based query form</td>
<td>E-mail</td>
</tr>
<tr>
<td>Nolo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Find/SVP</td>
<td>Business</td>
<td>Fee-based ($250 per question)</td>
<td>Findsvp.com</td>
<td>Business experts</td>
<td>Input query through a Web-based query form</td>
<td>E-mail, phone, fax, courier</td>
</tr>
<tr>
<td>Professional City</td>
<td>Law, accounting, marketing</td>
<td>Fee-based ($25 for the first 15 minutes and $18.75 per additional 15 minutes of cybrarian's time)</td>
<td>Professional City.com</td>
<td>Cybrarians</td>
<td>Input query through a Web-based query form</td>
<td>E-mail</td>
</tr>
</tbody>
</table>
the different types of services and some of their characteristics. It provides the following facts about the Web-based reference services listed there:

- the listed Web-based reference services are offered by dotcom companies;
- these services use the Web only for communication between the user and the system/answer provider, while the information service is provided by a human expert;
- while most services are free, some charge as high as $250 per question;
- in many cases services are provided by (self-proclaimed?) volunteer experts;
- there are some services offered in specific subject fields.

In addition to those in Table I, there are also some Web-based reference services where users need to conduct a search for a reference query. Such services provide free access to various online reference sources, and allow users either to select a specific source or conduct a search on a range, or all, of the reference sources. Examples of such services include the following:

- Internet Public Library (http://www.ipl.org).
- Infoplease (http://www.infoplease.com).
- Britannica (http://www.britannica.com).
- Internet Library for Librarians (http://www.itcompany.com/inforetriever/).
- ReferenceDesk (http://www.referencedesk.org/).
- Xrefer (http://www.xrefer.com/).

While most of these Web-based reference services are available free, some charge a small fee. For example, the Electric Library charges $79.95 for an entire year of unlimited access.

Janes et al. (2001) reported a study of 20 Web-based “expert services”. By asking 240 questions to 20 selected expert services, they noted that the sites gave verifiable answers to 69 per cent of factual questions. An interesting observation of this study was that the kind of reference interview that takes place in traditional reference service environments is non-existent in Web-based reference services. In fact, the sites did not come back for any discussion or clarification after receiving the initial queries. As a justification, the authors suggested that Web-based information services have been built mainly to answer factual questions, and therefore the experts concerned do not
need to go through the reference interview process. Nevertheless, the high rate of success of factual questions shows that more and more end-users will move towards these expert services rather than to libraries for answers to simple (ready reference) types of questions.

**Search engine services**

Although Ask Jeeves is basically a search engine, many researchers also see it as a Web-based information service because of two reasons:

1. unlike in other search and metasearch engines, users can ask a question in Ask Jeeves, and in many cases can get an answer right away; and

2. users can ask a question on a given topic, and Ask Jeeves comes up with a list of questions on the same or similar topics; the user can select any of the those predefined questions, and then Ask Jeeves provides answers to that.

This is an interesting service and may be considered as a useful model for digital library reference and information services. The Ask Jeeves site (http://www.AskJeeves.co.uk) maintains that:

Ask Jeeves is a unique question-answering system that allows users to ask questions in plain English, and then delivers the answer. As one of the most sophisticated navigation systems on the Internet, Ask Jeeves combines a unique natural language engine with a proprietary knowledgebase. Taken together, this mechanism processes the meaning and grammar of real questions in plain English; provides intelligent responses for user confirmation; links directly to relevant, high-quality answers; and, perhaps most exciting of all, becomes more intelligent as its knowledgebase expands with each question asked and each answer delivered.

The site further says that Ask Jeeves processes each query syntactically (to analyse the grammar) and semantically (to determine meaning), and then Ask Jeeves’ answer-processing engine provides the question template response (the list of questions that users see after they ask a question). When the user clicks on a response, the answer-processing engine retrieves the answer template that contains links to the answer locations. Thus, Ask Jeeves helps users select a query from a pre-defined set of queries on a given topic. However, there is a debate on whether the kind of service provided by Ask Jeeves matches reference services provided by libraries. Kresh (2000) reported an interesting finding: “One consortial system in northern California conducted an informal test of Ask Jeeves by sending 12 questions its libraries had answered. There were no trick questions, none were arcane, just questions typically received by those libraries. Jeeves was unable to answer any of the questions”. Nevertheless, the techniques and the technology used by Ask Jeeves may be very useful for introducing reference services in digital libraries.

Webhelp.com (http://www.webhelp.com) is another service that claims to offer “real-time search assistance with a real live expert – any time, day or
night”. About.com (http://www.about.com) is a service that shows a number of pre-defined categories related to a search topic given by the user. For example, a simple search on “e-commerce” not only produces a list of sites, but also provides a set of topics related to “e-commerce”, such as e-commerce definition, e-commerce security, advantages of e-commerce, e-commerce disadvantages, e-commerce statistics, etc.

Digital reference services and libraries
A number of libraries have now begun to offer Web-based reference services and a number of recent studies report the current practice of reference services provided by libraries. Although this is not an exhaustive review of all the Web-based reference services provided by libraries, the following section provides a quick overview of some that are currently available.

Digital reference services for the general public
Ask A Librarian (n.d.) is a Web-based reference service, primarily designed for UK residents, provided by a network of public libraries in the UK. The site says: “Mail us your factual question and we’ll send you an answer within two working days, if not before”. A user has to put the query through an enquiry page, which is automatically routed to one of the participating libraries, which receives it as an e-mail message. Within two days the library sends an e-mail message to the user with its response to the enquiry.

The British Library provides special reference services for business, patent, scientific, technical, medical and environmental information. These services range from answering simple questions to finding answers to complex questions involving online database searching, etc. While some of these services are free, for others users need to pay. For example, users can ask simple business questions using a form, and can expect an answer within ten working days. Similarly, users can send e-mails with simple environmental queries. A typical answer in such a case may include (British Library, n.d.):

- a list of bibliographies from a British Library catalogue literature search; and/or;
- a list of organisations to contact for more detailed information, and/or
- information retrieved from Internet sources.

Users are not charged for this kind of service. However, for complex queries users need to pay. For example, the British Library provides a Science Technology Medicine (STM) search service for which users need to pay at the following rate (British Library, n.d.):

- £82.00 per hour of staff time divided into 15-minute periods;
- plus costs of online searching;
- plus Value Added Tax (VAT).
As stated in the Web site (British Library, n.d.), most searches take at least 30 minutes and therefore cost £41.00 for staff time + costs for online searching + VAT. The site also mentions that the average list of references costs £80.00 for medical subjects, and between £100.00 and £150.00 for other subjects. Users are also charged a fee if the results are to be faxed. In order to obtain an STM search service, the user has to fill in a form that has three parts. In the first part users need to fill in their personal details, etc.; in the second part, users need to enter the query, and in the third part they need to specify the modes of delivery, payment, etc. In the second part of the form, users can enter the detail specification of the query using Boolean operators and specifying date, language, etc., as limiting factors.

**Digital reference services for users of academic libraries**

Academic libraries have also begun to offer Web-based reference services. Wasik (n.d.) and Wasik and Lankes (1999) discuss the value of digital reference and AskA services in the K-12 educational environment. They described how AskA services are built and maintained, and also explain how the service works in the classroom. Archer and Cast (1999) emphasised the importance of the personal touch in reference services and discussed how Web technology and the personal element of reference services can be combined to provide Web-based reference services. Tenopir (2001) and Tenopir and Ennis (2001) reported a survey of the current practices of digital library services in 70 academic libraries in the USA. These studies noted that the university libraries allow their patrons to pose reference questions in a variety of ways: 99 per cent offer e-mail reference, 96 per cent offer reference services by appointment, and 29 per cent of the libraries offer real-time virtual reference. Breeding (2001) briefly described the various customer relationship management (CRM) software, such as eGain (http://www.egain.com), LivePerson (http://www.liveperson.com) and WebLine (http://www.webline.com/products/web.htm), that are used by libraries for providing Web-based reference services. He also suggested that while the CRM software offer a great deal of sophistication to the virtual reference environment, they are more complex and expensive compared to lower-cost chat-based utilities that allow a managed two-way text conversation between the reference provider and the remote library user, thus replacing the need for a face-to-face communication.

Smith (2001) discussed the current technologies, such as chat and videoconferencing software used by libraries for providing digital reference services. Richardson et al. (2000) examined the information technology aspects and the key organisational issues involved in establishing an electronic reference desk service in a library. They also reviewed the usefulness of some electronic reference services. Breeding (2001) provided an overview of some methodologies and collaborations currently in use to extend assistance to
patrons in learning to use virtual resources and in finding the information they need.

**Digital library service by co-operative library systems**

In order to deal with the rising cost of reference sources and staff and also of the huge initial investment required for introducing new services in libraries, many have gone for a co-operative model of digital reference services. Oder and Weissman (2001) described the 24/7 Reference Project at Metropolitan Cooperative Library System, a consortium around Los Angeles, and the Bay Area Libraries Project at San Francisco that are co-operative projects for providing Web-based reference services to customers.

One of the advantages of Web-based reference services is that users can ask for such services from a remote location at any time of the day. However, Rogers (2001) presented an interesting report of a study analysing Web-based reference services in academic libraries in Illinois. This study shows a usage pattern that is quite similar to the timing of usage of traditional reference services in libraries, for example, 80 per cent of all usage occurred between 8 a.m. and 10 p.m.; evening use was higher than morning use, and the busiest hour was between 2 p.m. and 3 p.m.

**Current digital library research on personalised services**

Meyyappan *et al.* (2000) and Chowdhury and Chowdhury (2000) have recently reviewed the status of some digital libraries from around the globe. These reviews reveal that while the digital libraries have very good information access and retrieval mechanisms in place, there is almost no provision for quality reference and information services. However, these studies show that there is a trend towards the personalisation of digital library service, as in case of HEADLINE (http://www.headline.ac.uk) and GEMS (presently called iGEMS, a digital library developed at Nanyang Technological University in Singapore).

Some recent research efforts have been directed at building personalised digital library environments. These studies can be grouped into three categories: user needs assessment, personalised user interface and personalised digital library (Meyyappan *et al.*, 2001a), Marchionini and Komlodi (1998) reviewed literature on user interfaces, particularly in a digital library environment, and suggested the need for a task-based digital library. Zhao (1998) discussed the personal digital library (PDL) that acts as a front-end to other electronic library systems and manages the user’s information needs. Barry and Barbara (1999) discussed the creation of personal digital libraries. The North Carolina University Library (NCSU) developed MyLibrary @NCState (n.d.) as a portal application to the NCSU Library’s information resources. This portal allows user profiling to build a dynamic, customised gateway to both general and discipline specific resources. MyLibrary (Cohen
is a Cornell University Library initiative to provide personalised library services to their students, faculty and staff on the basis of a focus group study to gauge library users. The MyLibrary project has two components, MyLink and MyUpdate. MyLink allows patrons to save useful information resource links that they found themselves or those suggested by the librarians via a targeted notification in their personal space. Users can access this personal space from any place. MyUpdate periodically queries the online catalogue and notifies users, utilising a pre-defined user profile. HeadLine (Gambles, 2000) is one of the eLib Programme’s Phase 3 projects, developing a hybrid library system called the HeadLine Personal Information Environment (PIE). This PIE uses portal-type technology to present an information environment that is personalised to the user’s needs and allows user customisation. This project presents users with pages of resources relevant to their courses/department. PIE users are also provided with an “all resources” page that contains all the library resources. Users can create their own lists of resources on their own pages. Meyyappan et al. (2001a, b, c) reported on ongoing research at Nanyang Technological University in Singapore on the design and development of task-based digital library organisation and a digital work environment (DWE).

The major objective of these research projects is to allow users to personalise digital library access so that they can view their preferred resources, and hence information is automatically filtered. These projects aim to select automatically a subset of the DL collection based on the nature and characteristics of a user. The user then can add new resources, or delete recommended resources, to his or her page. The DWE project (Meyyappan et al., 2001a, b) is somewhat different from others in the sense that the main objective of the DWE is to provide filtered (based on the chosen task of a given user) access to the local digital library collections, to remote digital libraries, as well as to traditional libraries, and, most importantly, to the vast information resources on the university intranet. The system also allows users to create their own personal workspace to store selected information resources for future use.

Current digital library projects on digital reference services
Some research projects aimed at providing reference and information services as part of the digital library services have recently begun. The most prominent is the Collaborative Digital Reference Service (CDRS) (n.d.) project launched by the Library of Congress. Other projects are funded by the US Digital Library Initiative phase 2. Brief information about these projects is presented in the following sections.

CDRS
Rudner (2000) commented that “by applying the best of what libraries and librarians have to offer (structure and organisation, in-depth subject expertise,
and analog collections) to the labyrinthine universe of unstructured and unverified information on the Internet, we can begin to bridge the gulf that exists between providers and users of information”. With this end in view, the Library of Congress has launched the Collaborative Digital Reference Service (CDRS), the mission of which is to provide “professional reference service to users, anywhere any time, through an international, digital network of libraries”. The CDRS is a library to library network for asking and answering reference questions. It has three main components:

(1) Member profiles: that contain information on member strengths and features.

(2) Request Manager: a software for entering, routing and answering reference questions.


The system is designed to work in a very simple manner. A participating library sends a question to the system on behalf of a user. The system looks into the knowledgebase for a match. If the answer is found, a response is sent to the querying library by e-mail. If the question (and the corresponding answer) is not found in the knowledgebase, then it is passed on to the Request Manager, which forwards the query to an appropriate library selected from the profiles of the participating libraries and their collections, expertise etc. This library receives the query, finds an answer and sends it back to the results store, which then sends the response back to the asking library, and stores the query and the answer in the knowledgebase. Hence, the strength of the system lies in the knowledgebase and the Request Manager. Since June 2000 CDRS has been a “working” pilot with a current total of 220 active members who are routing real questions and answers. The next generation on CDRS is being co-developed by Library of Congress and OCLC. It will integrate local and global e-reference networks, and will be released as a full production service in June of 2002 (Kresh, 2000, 2001; Grotke, 2002).

**Automatic reference librarians for the World Wide Web**

Etioni and Weld (n.d.) reported on the project “Automatic Reference Librarians for the World Wide Web”, a DLI-2 funded project at the University of Washington. The central objective of this project, according to them, “is to create software agents that possess reference intelligence – a limited understanding of complex technical topics, but a very sophisticated understanding of how and where to find high-quality information on the World Wide Web”. Etioni and Weld reported the basic working principles of the proposed system, which works on a wrapper technology. The basic steps involved are:
the user asks a question;
• the Query Router assigns a topic to the query;
• the topic maps to a number of relevant wrappers;
• the parallel Web search module sends requests via wrappers to the sites;
• responses from the sites are obtained and sent to the fusion engine for collation;
• user gets the response.

The success of this system depends on the wrappers. Simply speaking, a wrapper is data that precede or frame the main data or a program that sets up another program so that it can run successfully. In the project the Automatic Reference Librarian program explores Web directories such as Yahoo! to find searchable sites. It queries each searchable site and obtains responses from them. The responses and other information about a given site are used to assign topics to that site. Thus, each searchable site gets a wrapper containing some assigned topics which are used for matching the topics of user queries.

**SIFTER**

SIFTER is an interdisciplinary research collaboration among faculty and students in the Computer and Information Science Department and the School of Library and Information Science at Indiana University. The broad aim of the project is to develop information agents that would perform a number of functions such as culling information from complex resources residing in diverse locations, and conducting analysis, synthesis and customisation according to the requirements of the user.

The specific objectives are (About Sifter, n.d.):
• dealing with heterogeneous information sources that may change over time;
• handling shifting requirements and interest of users with minimal human involvement;
• creating a robust agent architecture for reactive, proactive and collaborative filtering;
• supporting agent collaboration based on natural or artificial economic frameworks for multi-agent tasks.

**Virtual Reference Desk**

The Virtual Reference Desk (VRD) is a project sponsored by the US Department of Education, and its objective is “the advancement of digital reference and the successful creation and operation of human-mediated, Internet-based information services” (Virtual Reference Desk, n.d.). VRD
defines a digital reference service as Internet-based question-and-answer services that connect users with experts. The basic idea of VRD is that when a given user (a K-12 student, a teacher or a parent) asks a question that cannot be answered by a participating centre, it is forwarded to the VRD network for assistance. VRD services include the following (Virtual Reference Desk, n.d.):

- Collaborative AskA Service: a network of AskA services and volunteer information professionals.
- The Learning Center: a Web site for the K-12 community with curriculum-related Web sites, frequently asked questions, and other previously asked questions.
- AskA+ Locator: a searchable database of high-quality K-12 AskA services.

VRD also supports research on interoperability standards, metadata and other aspects of digital reference services.

*Tracking footprints through an information space: leveraging the document selections of expert problem solvers*

This is a DLI-2 project at the Oregon Health Sciences University and Oregon Graduate Institute. The goals of this project are “to understand how experts select information in a large and complex information space and to develop tools that assist them in this process” (Gorman *et al.*, 2001). The project aims to capture the trace of information used by experts in the health field, for example medical doctors. The basic idea is that as medical experts traverse the large and diverse collection of documents, they make explicit choices about which information resources to examine more carefully and which ones to ignore. These choices create a subset of documents relevant to a given problem and this subset of documents may be of interest to other users concerned with the same or similar problem. Thus, by capturing this information that is inherent in the selection of appropriate information resources for problem solving, this project aims to come up with models that may be useful for other users. The project proposed to develop (Tracking Footprints through an Information Space, n.d.):

- a trace that describes the path taken by an expert in solving his/her problem by using the available information resources;
- a précis of information about each document, its content, and the history of its use by experts; and
- navigation tools that would assist subsequent problem solvers using the collection by exploiting the knowledge inherent in the existing traces.

Though the project specifically deals with medical experts and looks into their problem-solving processes, the project team expects that the resulting
technology will be scalable, because the value of a set of précis and bundles within a collection should increase as the document collection grows and as more experts use the collection.

**Question answering systems**

A number of question answering systems (QA) are now being developed that aim to provide answers to natural language questions. Research in this area will have tremendous impact on the future of digital reference services. QA systems often use a variety of natural language processing (NLP) tools and techniques to get the correct answer from the source texts. Breck *et al.* (1999) reported a QA system that uses techniques from knowledge representation, information retrieval, and NLP. Research reported in the QA track of Text Retrieval Conferences (TREC) show quite encouraging results. The goal of research in the QA track is to “retrieve small snippets of text that contain the actual answer to a question rather than the document lists traditionally returned by text retrieval systems” (Voorhees, 1999). The basic technology involves analysing user queries in several stages. First, cue words/phrases like “who” (as in “who is the prime minister of Japan?”), “when” (as in “when did the Jurassic period end?”) are identified to guess what is needed; and then a small portion of the document collection is retrieved using standard text retrieval technology. This is followed by a shallow parsing of the returned documents for identifying the entities required for an answer. If no appropriate answer type is found then the best matching passage is retrieved. This approach works well as long as the query types recognised by the system have broad coverage, and the system can classify questions reasonably accurately (Voorhees, 1999). In TREC-8, the first QA track of TREC, the most accurate QA systems could answer more than two-thirds of the fact finding type of questions (e.g. “how many calories are there in a big Mac?”, “where is the Taj Mahal?”, etc.) correctly. In the second QA track (TREC-9), the best performing QA system, the Falcon system from Southern Methodist University, was able to answer 65 per cent of the questions (Voorhees, 2000). These results are quite impressive in a domain-independent question-answering environment. However, the questions were still simple in the first two QA tracks. In the future, more complex questions requiring answers from more than one document will be handled by QA track researchers. Such robust QA systems can be used to provide reference services in digital libraries.

**Evaluation of digital libraries and their services**

Evaluation data of digital libraries and their services are not yet widely available. In fact, very few evaluation studies of digital libraries have taken place so far. The most prominent publications reporting some digital library evaluation experiments appear in *Library Trends* (2000). Saracevic (2000) discusses a number of issues related to the methodology and criteria for
evaluation of digital libraries. Arguing that digital libraries were still at the stage of research and evolution, he commented that the best practice standards have yet to develop against which the performance levels are to be measured. Carter and Janes (2000) reported a study that analysed the log of over 3,000 reference questions asked of the Internet Public Library (IPL), which is one of the most widely used reference services on the Web. This study has identified different categories of reference questions, such as common questions, quick questions, regular questions and unanswered questions, which were handled by IPL. The study revealed that users seem to have difficulties in assigning subject categories to their questions, and also in determining whether they are factual or require detailed information searching.

Since the greatest part of reference and information services in the past has revolved around assisting users in their information search, and because of the fact that in a digital library environment end users need to search information resources all of which are online, it may be wise to take a quick look at some recent studies involving end-users’ search behaviour in the online and Web search environment. Pending any data from large-scale digital library evaluation experiments involving users, findings of these studies may help us assess the users’ online information search behaviour.

Gordon and Pathak (1999) compared the retrieval effectiveness of eight search engines: Yahoo!, Excite, Magellan, HotBot, AltaVista, Lycos, Infoseek and Open Text, and noted, among other things, that there were no statistical differences in the retrieval effectiveness among users for recall, though there were for precision. Although this study talks about the performance of the selected search engines, this observation is very important as far as the need for personalised help in digital library searching is concerned. It shows that the Web search tools differ in terms of search precision, and who can argue that in an online search environment of today, precision is a very important issue since no one will ever have time to sift through the hundreds of thousands of documents retrieved by an online search tool at any given instant. Jansen and Pooch (2001) made a comparison of user search patterns among three online search systems. Major findings of this study, as presented in Table II, show some interesting searching behaviours of users across the systems.

These observations show the rather poor state of end-user searches. For example, users used on an average two terms per Web search query, and only 8 per cent of the search queries contained Boolean operators. Similar observations have been made by another study conducted by Spink et al. (2001). They analysed over 1 million Web queries to discern how the public searches the Web and had some astonishing findings, as follows:

- a great majority of Web queries are short, not much modified, and very simple in structure;
Table II. 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>Number of queries per user per session</td>
<td>1.2</td>
<td>7.16</td>
<td>2.5</td>
</tr>
<tr>
<td>Number of terms per query</td>
<td>2</td>
<td>6.9</td>
<td>1.2</td>
</tr>
<tr>
<td>Number of relevant documents viewed per session</td>
<td>10 or less</td>
<td>Approx. 10</td>
<td>Less than 50</td>
</tr>
<tr>
<td>Number of queries containing advanced options</td>
<td>9%</td>
<td>9%</td>
<td>8%</td>
</tr>
<tr>
<td>Number of queries containing Boolean operators</td>
<td>8%</td>
<td>37%</td>
<td>1%</td>
</tr>
<tr>
<td>Number of queries improperly formatted</td>
<td>10%</td>
<td>17%</td>
<td>7.9%</td>
</tr>
</tbody>
</table>

- very few queries incorporate advanced search features, and when they do so, half of them are mistakes;
- despite gathering a large number of Web sites as answers to their queries, users do not view beyond the first or the second page of results;
- Web users are not much interested in relevance feedback.

Yet another recent study by Cooper (2001) reveals similar findings. Cooper 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. The findings of this study reveal that:

- the length of time research searchers spent at a session grew from six to ten minutes during the 16 month study period;
- during the study period, about 7.4 million database selections were made, out of which the catalogue database accounted for 32 per cent (about 2.4 million uses), Medlars for 22 per cent, magazine database for 10 per cent, periodicals database for 8 per cent, Inspec for 4 per cent and Current Contents for 4 per cent;
- the time users spent on each database also varied: about 3 minutes per search in the Medlars database, while 2.2 minutes for the catalogue database;
- users displayed about four to five citations for each search they performed;
- the length of time users spent displaying results was 30-40 seconds.

The observations of these studies are very useful for designing digital library services. One of the major problems of a digital search environment is the absence of human intermediaries who play a major role in traditional online search services. Appropriate online support for query formulation and modification may improve digital library search results. Again, since most
users do not look at more than a handful of records, appropriate mechanisms have to be developed for filtering retrieved output.

**Summary**

The discussions so far in this paper reveal some interesting facts that may be very useful in shaping the nature of personalised services in second generation digital libraries.

Although, as shown by Borgman (1999), the research definitions of digital libraries focus mainly on information access, retrieval, infrastructure, etc., there are some definitions that point out the need for personalised services. The most comprehensive connotation of digital library services appears in the editorial by Marchionini and Fox (1999) in the special issue on digital libraries of *Information Processing & Management* (1999). Cronin (1998) added an additional feature to the roles of reference and information services in the digital age proposed by Marchionini and Fox, and that is “to exhibit and facilitate digital literacy”.

Discussions on the current state of Web-based reference and information services revealed the fact that in most cases the Web is used only as a medium of communication, for example for sending the questions and answers. Sherman (2000) made a comparative study of three Web-based reference services: Ask Jeeves, Electric Library, and Information Please, and noted that:

- Ask Jeeves is useful for complex questions, and is a good choice for searchers that lack Boolean or other searching skills, because of its strong natural language parser and question-and-answer template structure.
- Electric Library is an excellent choice for a serious researcher in need of timely content from a wide array of otherwise unavailable sources.
- Information Please is an excellent tool for students and other researchers, as an authoritative source of facts and pointers for further investigation.

Janes *et al.* (2001) noted that the Web-based “expert services” were quite successful in providing answers to factual questions. Their conclusion has been that “the skills and knowledge of trained information service professionals may be better suited to detailed and source dependent questions in any event, and that the digital domain provides an environment rich with the potential for reference services to be greatly enhanced”.

Current research on personalised services in digital libraries reveals two different trends:

1. the creation of personalised work space for users where they can store information that they find useful; and
(2) automatic filtering of information is done based on user categories, user task, etc.

Coffman and Saxton (1999) found that a networked reference service based on the call centre model could reduce the reference staff cost significantly. However, they caution that though the call centre model has been quite successful in answering customer questions in the business environment, it may not be quite appropriate for answering reference questions since they may come from anyone on any topic, asked in any form.

The evaluation studies reported in this paper reveal that:

- users find it difficult to categorise their reference questions;
- 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 the retrieved items;
- very few queries contain advanced search features.

A brief review of some digital library research projects focusing on the digital reference and information services reveals that:

- some projects aim to use the agent technology for identification, filtering and categorisation of information;
- collaboration among libraries and various institutions has been considered as a measure to cut the cost and achieve higher performance levels;
- one project is trying to create models of expert users’ information problem solving techniques so as to create some generalised models to be useful for other users.

**Conclusion**

Lankes *et al.* (2000) have pointed out that “the reference librarian in the new millennium will need the ability to read the situation a user is in and find the right information for that situation”. Considering this view and also to keep in pace with the rapid appearances and developments of Web-based reference and information services provided by non-library organisations, many library and information science professionals have now turned their attention to the provision of e-reference services. Oder and Weissman (2001) suggested that “the year 2000 brought the advent of live reference. Several libraries, especially academic ones, have used or adapted chat or commercial call centre software to communicate with surfers in real time and send Web resources to their browser.”
Indeed, this is a time when digital library researchers should think of the best ways and means to make the optimum use of the technology and experience and expertise of human intermediaries in improving digital libraries from mere access centres to information service providers. Many recent projects have taken the approach to building software agents to do much of the work of reference librarians. While this is a useful approach, and such programs will reduce the workload of human experts, it is doubtful whether digital libraries can totally replace human experts and personalised services. So far as the information filtering aspect of the personalised services is concerned, the personal digital library approach of automatic filtering based on user nature, and the DWE approach (Meyyappan, 2001a, b) of information filtering based on user categories and tasks, seem to be quite promising.

Recent research on online search behaviour of users re-emphasises the need for user training and assistance in online searching. This is in contrast to the views of Arms (2000), discussed earlier in this paper, who proposed that information retrieval tasks can now be successfully performed by end users without the assistance of reference librarians. User training, or information literacy programmes, may be conducted regularly and sophisticated online tutorial programmes using digital video technology may be developed to assist users. For assistance in conducting complex subject searches, the current practice of e-mail communications, online chat, videoconferencing, etc. are useful and may be deployed depending on the nature, number, location, available IT resources etc. of the end users.

The approach taken by Ask Jeeves in the use of templates for storing queries, and the knowledgebase for storing the queries and their answers may be adopted by digital libraries. The strength of Ask Jeeves is the knowledgebase which, according to the site, “contains links to more than 7 million answers, which contain information about the most frequently asked questions on the Internet”. Library and information professionals are trained in organising information and a new research may be conducted for finding the best way to organise reference queries and answers on various topics. A facet analysis technique may be used in formulating different templates around a given search topic and they may be stored in a knowledge base. A knowledge base, developed on such an approach, will grow fast initially, but after a certain time the growth rate will be minimum. Only when new aspects of a given topic appear in the research and literature will there be a new impetus.

Co-operation in providing ready reference and subject-specific reference services may be a useful cost-cutting measure. Indeed, the high cost of staff in libraries has been a major problem. However, since digital libraries can be accessed and queries easily dealt with from remote locations, co-operative mechanisms based on subject specialisation and other criteria may be considered.
Specialised search and other forms of assistance need not always be free. In fact, some services mentioned earlier in this paper charge hefty fees and this is perfectly reasonable, especially where specialised and reliable searches can save end-users’ valuable time. Payments for such services may not always necessarily come from the end users. They may be recovered from research projects (in an academic institution), business projects (in a business organisation) and so on. Holmes-Wong (1999) suggested that, by combining Internet technology and electronic commerce, libraries may be able to offer customisable electronic reference services to users. Nevertheless, new research for testing e-commerce models for digital library reference services needs to be undertaken.

To sum up, this study re-emphasises the need for, and provision of, personalised information services in the next generation digital libraries. The current research projects aiming at finding the best means of providing such services through digital libraries are trying a variety of techniques and technologies, including:

- information filtering mechanisms based on users’ personal characteristics and tasks;
- agent technology for automatic information discovery, organisation, filtering and retrieval based on user nature, need, etc.;
- co-operative models for sharing the resources and specialisations – both the content and staff;
- Internet technologies such as Internet chat, call centre management software, video conferencing, etc.

These research projects will definitely bring some useful models and approaches that the second generation digital libraries may adopt. Further research may be carried out in a number of areas, including:

- building appropriate packages for end-user training involving the best practices for digital library use suitable for specific user needs, characteristics, etc.;
- testing the query template and knowledge-based approaches in a digital library environment (approaches somewhat similar to those of About.com and Ask Jeeves, for example);
- testing e-commerce models for handling the payment and cost management aspects of sophisticated and resource-intensive personalised services;
- tracing the techniques of expert information problem-solving processes, and building models for other users in a given domain; and
• testing traditional library classification tools and facet analysis techniques for creating question templates and knowledge bases for storing answers in specific subject fields.

References


This article has been cited by: