Enabling the Collaborative Construction and Reuse of Knowledge through a Virtual Reference Environment.

J. Alfredo Sánchez, Alberto J. García, Carlos Proal, Lourdes Fernández
Library and Center for Research for Information and Automation Technologies of the
Universidad de las Américas-Puebla, Mexico
{alfredo, alberto, carlos, lulu}@mail.udlap.mx

Abstract

This paper analyzes some of the factors that prevent knowledge construction and reuse in traditional library settings and describes *VRef*, an environment designed to provide users with the functionality of digital libraries in the areas of collaboration and social construction of knowledge. VRef makes it possible for users to post their information needs to a virtual space accessible to all system users. Information needs are addressed by reference librarians or by automated agents that use knowledge about problems solved successfully in the past. Whereas reference librarians have pre-assigned roles in this interaction, other users may volunteer their expertise and participate in the solution of open questions. The interaction among users playing different roles is recorded so successful responses form a valuable knowledge repository that can be reused and enriched by users and their agents. VRef facilitates the transition from the traditional reference desk on to the more dynamic role of a *digital reference librarian.*

Keywords: *virtual groups, collaborative workspaces, digital libraries, shared spaces, agents, hybrid physical/virtual environments.*

1. Introduction.

A recent report estimates that as much as two exabytes (billion gigabytes, or 10^{18} bytes) of unique information are produced every year in the world on all kinds of media [4]. According to the same study, only 0.003% of the generated information is comprised by printed documents, whereas the digital medium practically has become universal for the purposes of information storage and dissemination. Organizing, providing access to and transforming all existing information resources into meaningful knowledge for specific users clearly are tasks of fundamental importance and enormous proportions.

For centuries, libraries have played a key role in defining and implementing organization schemes as well as providing access for users to traditional analog materials, such as books, serial publications, microfilm and audio and video recordings. Librarians developed highly specialized skills and regularly enquire and elicit user requirements and fulfill information needs using resources at their libraries or through inter-library loan mechanisms.

As media for information storage and dissemination continue to evolve, it has become clear that traditional organization schemes and librarian skills neither map well into the digital realm nor they take full advantage of the new medium's potential. The entities now commonly referred to as "digital libraries" are not really libraries in the usual sense. In order to convey some of the basic functionality provided by digital libraries in terms of a well-known concept, researchers have resorted to the abstractions of *content* and *services* that generally describe physical libraries. But this metaphor does not even suggest the major differences that the introduction of a new medium entails. In addition to the new substrate, digital libraries comprise new information genres, allow for multiple novel organization schemes, and make diverse browsing and searching mechanisms possible. They also require entirely new skills for managing contents and providing services to users. What is perhaps more important is that digital libraries enable completely new ways for users to collaborate and participate in the social construction of knowledge.

This paper analyzes some of the factors that prevent knowledge construction and reuse in traditional library settings and describes *VRef*, an environment designed to provide users with the functionality of digital libraries in the areas of collaboration and social construction of knowledge. VRef makes it possible for users to post their information needs (typically addressed by reference librarians) to a virtual space accessible to all system users. Information needs are addressed by the staff of the digital library or by automated agents that use knowledge about problems solved successfully in the past (by human assistants or by other automated agents). Whereas reference librarians have pre-assigned roles in this interaction, other users may volunteer their expertise to participate in solving open questions. The interaction among users playing different roles is recorded so successful responses form a valuable knowledge repository for use by users and their agents. VRef facilitates the transition from the traditional reference desk on to the more dynamic role of a *digital reference librarian.*

The paper is organized as follows: Section 2 provides an account of the process most commonly followed at the traditional library's reference desk, which was a key motivation for the main development we describe here. Section 3 discusses the opportunities posed by the introduction of the digital medium into an environment that supports scholarly work. Next, we refer in Section 4 to the overall initiative that frames our digital library efforts and VRef in particular. The design of the proposed virtual reference environment is discussed in Section 5, whereas Section 6 presents our initial prototypes and user testing results. The relationships between VRef and other current work in collaborative knowledge construction and reuse are discussed in Section 7. Finally, the status and directions of our project are provided in Section 8.

2. Knowledge construction (and destruction) at the reference desk

Reference librarians perform an extremely important job that makes it possible for users to make the most of the resources available at their library. They accumulate knowledge about those resources and are able to direct users to books, articles and other materials which address their information needs. Even though on-line catalogs and various automated searching mechanisms have become widespread, the expertise and common sense of reference librarians still represent some of the most valuable assets a library has to offer to its patrons in need of timely and precise information.

Typically, several reference librarians serve a community of users with some degree of specialization as for the areas each librarian covers. By its very nature, the job of a reference librarian is one that requires continuous construction of knowledge about the resources at hand. However, it is not possible for academic and research libraries to have a reference librarian for every user's specific area of interest, so user needs must be satisfied with the resources available. Although some knowledge construction in these cases occurs as driven by user requirements, and results improve as librarians begin to know their users, knowledge is not recorded and is therefore unlikely to be shared and reused. Thus, every time a librarian receives a user request which has been successfully satisfied in the past by another librarian, all the work performed previously is repeated.

This problem can be particularly severe for the case of highly variable user communities (such as university libraries). As the subtitle suggests, at the traditional library's reference desk knowledge is constructed, not very easily reused, and eventually destroyed.

3. Challenges and opportunities for a virtual reference environment

A digital library is a virtual space that enables scholarly work, which nowadays regularly involves collaboration among potentially remote users. The concept of digital library comprises rich repositories of digital information and a wide range of digital services and interfaces or environments at the users' disposal. Digital library researchers recognize that human skills and human assistance are preferable for many users as they conduct information-intensive tasks. Technology should be thought of as the medium supporting the interaction among users engaged in scholarly activities.

Although users of a digital library will be provided with richer mechanisms for searching and browsing information resources, knowledgeable human assistants providing reference services in the digital realm still will be of great help. Moreover, digital libraries will coexist with their traditional counterparts for quite some time, thus making it desirable for reference librarians to develop new skills in order to be able to respond to the needs of users in the new setting.

In the context of a mixed traditional/digital library, reference librarians may direct users to three major kinds of resources: physical materials in the traditional library (books, journal articles, etc.), digital in the digital library contents (digital documents, software, etc.), and digital contents outside the digital library (web pages, ftp sites, etc.). Integrating resources so librarians and patrons can use them seamlessly is a challenge for researchers.

Perhaps the most important opportunity posed by the integration of traditional and digital libraries is the creation of spaces in which knowledge is constructed (and preserved) collaboratively by the participants of dynamic, distributed research communities. Not only can all the interactions among co-located and remote participants be recorded, but they can also be made available for all users so newcomers can easily move through paths previously traversed by others, experts can review, annotate and improve existing knowledge, and automated agents can assist all participants with their research. This is essentially the goal of the VRef environment.

In order to provide a better view of the development of VRef, we briefly describe next the overall digital libraries initiative that frames our project.

4. Digital libraries context

Our group has developed a system architecture for a digital library that addresses the needs for information management, communication and collaboration among a highly distributed community of users [3]. We aim to build both a practical, functional digital library and a testbed for research of open issues in the field, including aspects of collaborative work, personalized and group interfaces, annotation facilities and information visualization.

The realization of our digital library has called for research and development efforts in three major areas: building content, designing components to provide general infrastructure (such as information retrieval services, multimedia management and agent essentials) and user interfaces and environments. With regard to digital contents, we have made progress in the construction of collections of digital theses, special collections from our library and university publications. Our advances in the area of general infrastructure include a distributed framework that integrates services and interfaces, with implementations using KQML, CORBA and Jini [2] as well as components implementing various popular information retrieval mechanisms. Finally, in the area of user interfaces and collaboration environments, our work includes a visualization tool for large information spaces organized hierarchically [1], an environment for collaborative revision and annotation of digital theses, and recommendation facilities for library users.

One distinctive element of our work in digital libraries is our view of information and collaboration spaces as comprising both the digital and physical realms. We believe the needs posed by scholarly activities can be best satisfied if traditional resources as well as the novel digital facilities are accessible to users at any given time in a seamless fashion. In this regard, we have been integrating traditional on-line catalogs and user services (such as lending materials, selective information dissemination and other notification services) into our new developments.

The VRef project described in this paper aims to go further along these lines: not only are traditional and digital information resources integrated into a virtual reference environment, but also VRef fosters the incorporation of reference librarians into the wider collaboration community enabled by digital libraries.

5. The design of VRef

Figure 1 illustrates the major components and participants of our virtual reference environment: a shared virtual space, users playing various roles, automated agents assisting users, a knowledge repository, digital collections accessible locally or via the global network and physical collections pointed to by on-line catalogs. Each of these components and their inter-relationships are discussed in this section.

5.1 Shared space, user roles and knowledge base

Two basic roles exist in a typical reference desk setting: *regular user* and *reference librarian.* Regular users require assistance to find specific information in the library as part of their research work. Reference librarians are library personnel specifically hired to assist users in this task.

VRef has been designed to provide a virtual area where users can meet with reference librarians and other users and collaborate in activities requiring information in digital

or physical collections. Users can access VRef from their workplace anywhere in the global network. As users enter this shared space they can identify immediately which other users are logged in and what roles they play. If users so desire, actual photographs may be displayed associated with each participant in the system.
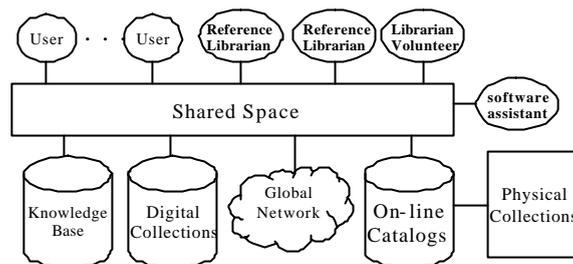

Figure 1. Major components and participants of VRef.

5.2 Collaborative knowledge construction

In VRef, users can post information requests for all the system participants to see, additionally indicating the subject or knowledge area in which the information required lies. New requests are immediately assigned a *pending* status. Depending on their workload, reference librarians may provide assistance practically in a synchronous way. Typically, however, requests will not only be posted during working hours (at least not for librarians, as users may be logged in from different time zones). Every time users log in to VRef they are presented with currently pending requests along with any answers already offered: for regular users these will be any requests they have posted, whereas for reference librarians pending requests will include those that best match their area of specialty, as registered in a user profile. In both cases users may opt to see all (other) requests sorted by the criterion of their choice (e.g. by user, by date or by topic).

VRef provides for the various specific types of answers that can be offered for information requests. For references to books, journals or other materials in the physical library answers include basic data from on-line catalogs such as title, authors, publisher, date of publication and location within the library. For digital materials both within or outside the digital library, users are provided with addresses that allow them to examine the suggested reading (or viewing). In all cases, answers may include comments regarding the pertinence, availability, or any other aspects of the suggested materials.

As can be imagined, some requests are amenable to be satisfied by users who do not formally work at the library. In VRef, provisions are included so anyone in the system can offer answers and comments to pending requests. In fact, regular users who would like to make themselves available to assist others in specific subject areas on a volunteer basis can register as Librarian Volunteers an can thus be identified when users enter VRef.

Upon receiving an answer, users who posted requests may reply to refine their initial request or may just indicate they have received answers to their satisfaction. In this case the request's status is changed to *successful*. If a request takes too long to be satisfied (according to each user's restrictions), its status is changed to *failed*.

5.3 Knowledge preservation and reuse

All interactions among VRef participants are recorded and form a valuable knowledge repository which is always available to all users. In this way, regular users may check the knowledge base including successful and failed previous requests when librarians are not available for assisting them. Similarly, librarians can use recorded knowledge when faced with recurrent requests or requests in areas outside their specialty.

The benefits of preserving this knowledge that has been constructed collaboratively can be further increased by mechanisms that facilitate its reuse. The design of VRef includes automated agents which also have access to the shared space, the knowledge base and the various collections. Agents analyze pending requests and (using similarity measures from the field of Information Retrieval) find information resources that may satisfy specific requests. Agent decisions are based on requests answered successfully in the past, information available in on-line catalogs and the actual contents of materials in digital collections.

Figure 2 shows a diagram representing a VRef scenario and provides a summary of the players and functionality considered in its design. In this scenario, three regular users (U1-U3), three reference librarians (RL1-RL3), one librarian volunteer (LV1) and one automated agent (AG) are participating in an ongoing VRef session. U1 has posed a question for which RL1, LV1 and AG have provided answers. U1 has indicated that the answers satisfy the request, which is marked as successful. U2 has made a request (Q2) for which only one answer has been provided (by LV1). Q2 is still considered pending, as user U2 may not have checked the suggested solution. U3 has made two requests (Q3 and Q4). A reference librarian (RL3) and the automated agent (AG) offered answers but they were not helpful or were not provided on time, thus making Q3 a failed request. On the other hand, two answers from RL2 and one from RL3 provided for Q4 did help the user and Q4 has been marked as successful.
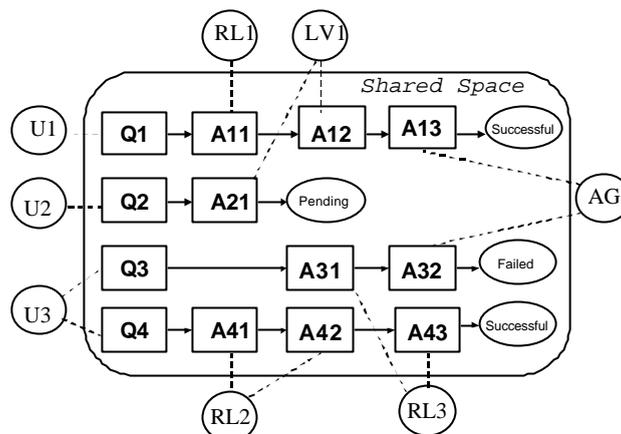

Figure 2. A typical VRef scenario (U_i = regular users; RL_i = Reference Librarians; LV_i = Librarian Volunteers; AG = Automated Agent; Q_i = Requests; A_{ij} = Answers)

6. VRef implementation and user testing results.

In order to explore various alternative representations and platforms for implementation, the design of VRef and the underlying data model were given to four programmer teams, each of which produced an initial prototype. Prototypes explored the use of various graphical layouts, menu arrangements and programming tools (including Java applets, interface toolkits, servlets and database drivers). Tests were conducted to receive early feedback from users mainly regarding system functionality and performance. For several years now, our Digital Libraries Group has included people from both an academic library and a research center in information technologies. Among other benefits, this has made it easier for us to get library personnel, including reference librarians, involved in prototype testing and participatory system design.

Considering the results of initial tests, we selected one of the prototypes to serve as the basis for further development. We now have a stable, multi-platform implementation of VRef which is based on Java servlets and a relational data model accessed via the JDBC interface. In this section we present the salient features of this implementation and discuss the results of various tests conducted with a real (though relatively small) user community.

Figure 3. Tasks, four users, and pending requests in VRef.

6.1 VRef functionality

VRef is accessible via a web browser from anywhere in the network. We have produced Spanish and English versions of the interface. For consistency, the English version is used in the examples included in this paper, though our experiences with real users have are based on the Spanish version. Figure 3 illustrates a session in our virtual reference environment. As can be observed, the user interface consists of three main areas: a list of tasks on the upper panel, a list of current users on the lower left panel, and a list of pending requests on the larger right panel.

The tasks listed in the upper panel vary according to the user's role. In the figure, from left to right the user may choose from posting a request (asking a question), returning to the list of pending requests, check the repository of requests solved successfully in the past, see listings of registered librarian volunteers and users who regularly post requests, review requests that have been left unfulfilled, and various configuration and administration options.

The figure shows four users who are currently logged in and their roles are identified by a small icon next to their names. In this case, two reference librarians, two regular users and no librarian volunteers are logged in. By clicking on their names, a user can obtain additional information about other users in the system, including photographs if available. The list of users is periodically refreshed by VRef.

The list of pending requests is sorted by topic, but the user may change this ordering by just clicking on the desired criterion. The requests shown have been posted by users not

currently logged in and their subject area is identified by a two-character abbreviation used at the library to refer to each of the majors offered by the university (in this case: architecture, physics/math, electrical engineering, music and international relations). By default, users are presented with a list of their personal pending requests, though they may view (and hopefully help address) all other requests. A discussion of non-obvious functionality and its implementation is presented next.
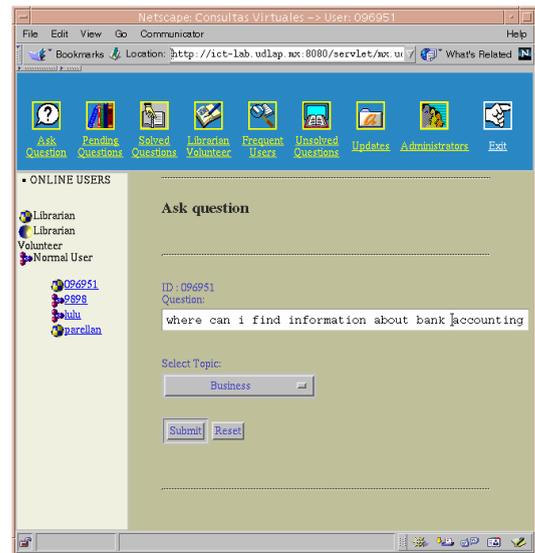

Figure 4. A simplified form to post requests to a shared space

Posting requests. Our initial approach for allowing users to post information requests is illustrated in Figure 4. Users enter their request in plain natural language and select one subject or knowledge area out of a predefined list. User testing indicated that encouraging users to provide as much specific information as possible from the beginning would shorten the process of eliciting their requirements with more precision. In a traditional reference desk setting, this is accomplished through an interview that helps in delimiting user needs. Our simplistic initial approach produced a series of questions and answers before the actual user request could be addressed. An evolution of this interface already in use provides for optional (though strongly recommended) input fields suggested by reference librarians, including publication dates, knowledge sub-areas, institution names, preferred languages and geographical references.

Offering assistance. As shown in Figure 3, with each pending request there is an option for librarians (or any user in VRef) to suggest information sources that could be of help. Users may choose from specific types of answers: books, serials, journal papers, materials in the digital collections and web sites references. Different templates are presented for each type of answer and multiple answers (of

diverse types) can be offered for a given request. Figure 5 (a) shows a template used to suggest books related to a given request. Note that free-format comments can be provided in addition to bibliographic data linking to an on-line catalog.

Receiving and rating assistance. When viewing pending requests (Figure 3), the last column of this listing shows links to any answers that have been already provided, ordered by type of answer. Figure 5 (b) shows the current interface for viewing suggested information resources. Once users have checked a suggested material, they are asked to provide feedback indicating whether their request has been fully satisfied. When they do so, their request is removed from the list of pending questions but is made available for later reuse through the list of successfully answered requests. If a request is not satisfied by a deadline (VRef's initial configuration sets the default deadline to 30 days from the date a request is posted), the request is removed and made accessible only through the failed requests option.
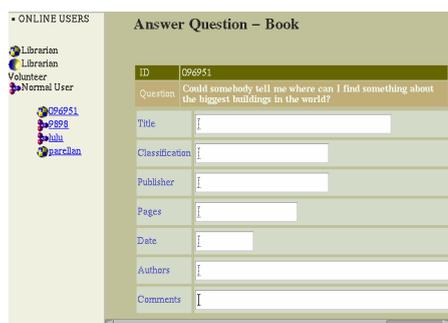
(a)
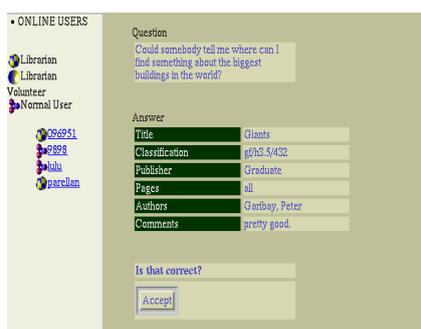
(b)

Figure 5. Offering assistance and reviewing answers in VRef

New role for the librarian. In the VRef setting, knowledge is constantly added to a permanently available database. New users may sign in and existing users may sign up to be librarian volunteers. Just like in the physical setting librarians are responsible for managing knowledge in analog sources, they become knowledge managers in the digital realm. VRef provides for an *Administrator* role,

which is granted only to users who are responsible for monitoring and "fine-tuning" the system. Users may see a list of administrators in case they need to contact one (Figure 6 (a)). Figure 6 (b) shows the interface currently available to administrators, which includes options to manage users, their roles, subject areas, and librarian-subject relationships.
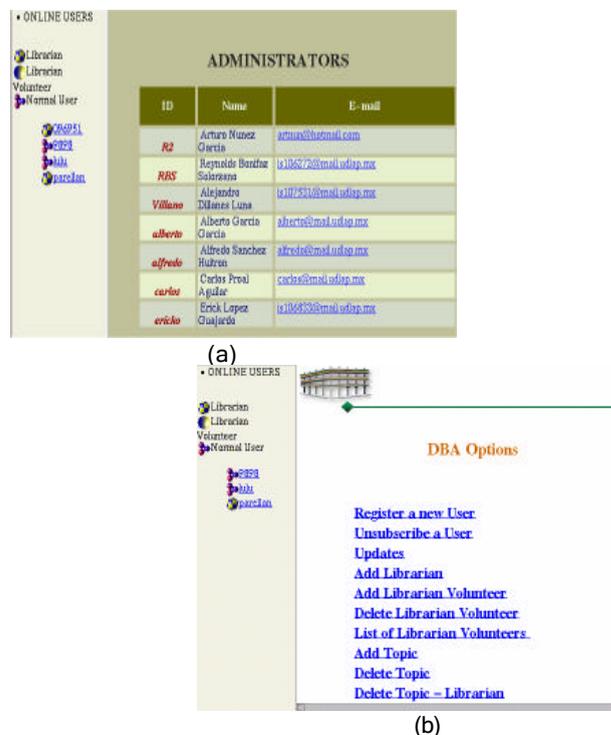

(a)

(b)

Figure 6. The librarian as knowledge manager.
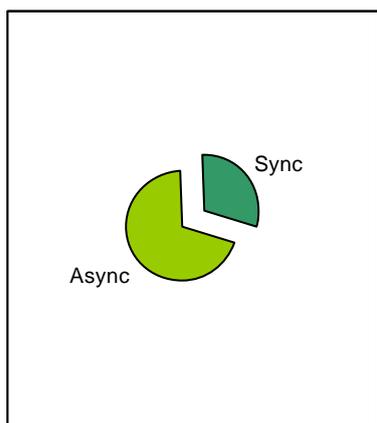
6.2 Results of VRef testing

As of this writing, VRef prototypes have been under testing for about six months and its current version has been in use for two months. For the time being its availability has been limited to our library and its user community. Our library's holdings include over 800 thousand documents in its physical collections and over six million documents in digital collections, mostly from external information providers. We expect to release a new version of VRef within two months (which will be open for the general public) so within six more months we will be able to share experiences recorded for about a year. The main users now include five reference librarians, four users who are librarian volunteers, and about 100 regular users we registered including faculty members, other library personnel and students. The entire local user community comprises over 7,000 members.

During a usability test phase, we explicitly surveyed our users regarding VRef's interface. User comments referred to VRef's similarity with popular web-based email and bulletin board systems, so most of them found VRef's

interface easy to use, as illustrated in Figure 7 (a). Users did comment positively on the personalization of the interface, as by default only relevant requests and their answers are displayed by VRef. System performance has also been acceptable, with average response times under two seconds for the most used system options.
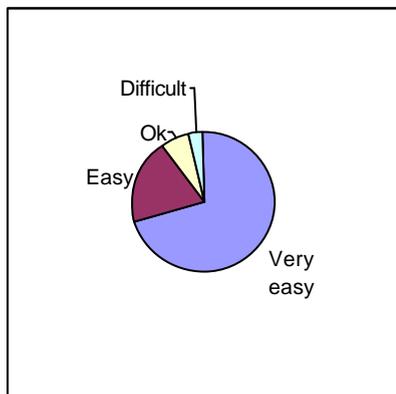
(a)

(b)

Figure 7. Some results from VRef testing

Interestingly, we did not expect much synchronous interaction to take place in VRef. However, as shown in Figure 7 (b), in the current stage a significant portion of users have tended to post requests only when reference librarians are present in the system and expect to receive immediate attention (which they usually do, given that the number of concurrent users has not been high yet).

Users do review the existing knowledge base and we already have observed some knowledge reuse. This has occurred for students in the same class in search of the same information. Instead of issuing the same request multiple times, when users log in and realize that someone has already asked about what they need, they just wait and examine the knowledge base when the request no longer appears as pending.

A natural problem we faced in the beginning was some resistance offered by our librarians to explore VRef as a means to provide services to the library's user community. As we incorporated their valuable feedback into VRef's design, we have observed increased participation and collaboration.

Agent participation. We have implemented an agent that extracts keywords from user requests and finds potentially helpful answers by using the vector space model [10] to measure similarities with existing responses in the knowledge base. Given the relatively short time VRef has been operating and the small size of its knowledge base, we have not included yet agent participation in the current version of the system. We expect this agent to be included in the next version of VRef with its own distinctive representation in the user interface. We expect to be able to report results from this development by the end of the summer. Agents based on other mechanisms are on schedule for later inclusion in VRef.

7. Related work

There is much ongoing work that is related to VRef's goals, involving researchers in fields such as CSCW, digital libraries, library science and knowledge management. In this section we discuss briefly some salient projects and their relationships with VRef.

The Collaborative Digital Reference Service (CDRS) is a large project launched by the US Library of Congress to form a network of libraries and particularly their reference areas to provide assistance to users around the world. As described by [6], CDRS proponents scheduled three major pilot phases, including: (1) the standardization of data elements required to describe library profiles and information requests; (2) training librarians and defining a governing board, and (3) defining an archive of questions and answers. In addition to scale and organizational issues, VRef's emphasis on digital collections and their integration with other materials, a bottom-up approach to facilitate collaboration (reference librarians first, library management afterwards), and rapid prototyping and refinement of our design distinguish our approach.

A project aimed to provide a large college community with a 24-hour help desk and the creation of a knowledge repository is described in [5]. They focus on computing services and do not consider specific provisions for the integration of digital and physical materials.

The work of Van de Sompel and Hochstenbach [11] on the SFX system for dynamic linking of highly heterogeneous sources presents an interesting approach to integrating information for users in a seamless fashion. In fact, some of the experiments conducted with SFX have involved commercial information providers which are popular among librarians. Whereas SFX emphasizes automatic link computation, VRef explores collaboration and human

participation in the construction of a useful, practical knowledge base. Issues considered by SFX should nurture our work in later stages.

Although richer and livelier interfaces for collaborative spaces were tempting in our initial design of VRef, we settled in our current version for a plain alternative which allowed for reasonable performance on the various system platforms available to our target users. Personalizable collaborative environments, as described for example by [9], are being considered as our software evolves.

There is also formal work researching the ways users go about finding information and information intermediaries elicit information needs (see, for example, [8] and [7]). We plan to consider their findings in refining our future work.

8. Ongoing and future work

VRef is a robust component of our digital library project. Its deployment is involving a growing community of users engaged in scholarly work and its benefits are just starting to show. As both the user community and the underlying knowledge base grow, so will the opportunities for collaboration and knowledge reuse. We plan to incorporate user feedback in our next version and continue to study and improve VRef's collaboration facilities.

As a new school semester starts in the fall, we look forward to two events: the appearance of recurring requests and the introduction of automated agents to assist users. In both cases we expect knowledge reuse to play an important role in satisfying information needs.

We plan to enhance VRef's functionality to move from a knowledge gathering and reuse mechanism to the facilitation of knowledge management for librarians and users in general.

Acknowledgments

References

[1] Proal Aguilar, C., Sánchez, J. A., Fernández, L. 2000. UVA: 3D representations for visualizing digital collections. In *Proceedings of the 3rd International Conference on Visual Computing* (Visual 2000, Mexico City, Sep.), 185-192.

[2] Barceinas, A., Sánchez, J. A., and Schnase, J. L. 1998. MICK: A KQML inter-agent communication framework in a digital library. In *Memorias del Simposium Internacional de Computación* (CIC, Mexico City, Nov.), 66-79.

[3] Sánchez, J. A. and Leggett, J. A. 1997. Agent services for users of digital libraries. *Journal of Networks and Computer Applications 21*, 1, 45-58.

[4] Lyman, Peter and Hal R. Varian, "How Much Information",2000.
Retrieved from:
http://www.sims.berkeley.edu/how-much-info.

[5] Graham, J., and Hart, B. 2000. Knowledge integration with a 24-hour help desk. In *Proceedings of the ACM Conference on User services: Building the future* (Richmond, Va., Oct.), 92-95.

[6] Kresh, D. N. 2000. Offering high-quality reference service on the web: The Collaborative Digital Reference System (CDRS). *D-Lib Magazine 6*, 6 (June).

[7] Nordlie, R. 1999. "User revealment"--a comparison of initial queries and ensuing question development in online searching and in human reference interactions. In *Proceedings of the 22nd ACM SIGIR Conference on Research and Development in Information Retrieval* (Berkeley, Calif., Aug.), 11-18.

[8] O'Day, V. L., and Jeffries, R. 2000. Orienteering in an information landscape: how information seekers go from here to there. Proceedings of the Conference on Human Factors and Computing Systems (CHI 2000, Amsterdam), 438-445.

[9] Portugal, R., Guerrero, L., and Fuller, D. 1999. A customizable collaborative virtual environment on the web. *Proceedings of the String Processing and Information Retrieval Symposium & International Workshop on Groupware* (Cancun, Mexico, Sept.), 328-335.

[10] Salton, G. 1989. Automatic Text Processing. Addison-Wesley, Reading, Mass.

[11] Van de Sompel, H., and Hochstenbach, P. 1999. Reference linking in a hybrid library environment. *D-Lib Magazine 5*, 4 (April).