The National Archives of Singapore reference enquiry database

Goh, Dion Hoe-Lian; Chan, Roy

2003

Goh, D., & Chan, R. (2003). The National Archives of Singapore reference enquiry database. The Electronic Library, 21(4), 316-321.

<https://hdl.handle.net/10356/91585>

<https://doi.org/10.1108/02640470310491540>

Goh, D.H., and Chan, R. (2003). The National Archives of Singapore reference enquiry database. *The Electronic Library*, 21(4), 316-321.

The National Archives of Singapore Reference Enquiry Database

Authors

Dion H. Goh

Division of Information Studies
School of Communication and Information
Nanyang Technological University
31 Nanyang Link
Singapore 637718
Singapore

Telephone: +65 6790 6290
Fax: +65 6794 0096
E-mail: ashlgoh@ntu.edu.sg

Dion Goh is an assistant professor at the School of Communication and Information, Nanyang Technological University, Singapore

Roy L. Chan

Division of Information Studies
School of Communication and Information
Nanyang Technological University
31 Nanyang Link
Singapore 637718
Singapore

E-mail: roycly@yahoo.com

Roy Chan is an Information Studies graduate from Nanyang Technological University, Singapore

The National Archives of Singapore Reference Enquiry Database

Keywords

Archives, Reference services, Information retrieval, World Wide Web

Abstract

The National Archives of Singapore offers reference services to members of the public on questions related to Singapore's history and heritage. However, with only two reference officers stationed at the reference helpdesk at any one time, users may often have to queue to await servicing while the officers are busy attending to queries, telephone calls or retrieving physical documents. The Reference Enquiry Database was thus conceived to provide Web-based access to ready reference information on frequently asked subjects by members of the public, thereby reducing waiting times for users, easing the workload of references officers and allowing them more time to attend to more complex information needs. This paper provides the background and rationale for the system as well as discusses design and implementation issues.

Introduction

The National Archives of Singapore (NAS) is a Singapore government organization that is tasked with the permanent preservation of the country's records which are of historical importance and significance (Siam, 2001). Established in 1968, it is responsible for acquiring, collecting, preserving, managing, and disseminating information about the country's history, culture, arts, and heritage.

The NAS identifies, selects, and acquires records of historical interest from local and overseas individuals and organizations. Intrinsically valuable records are preserved and conserved in their original form in environmentally controlled repositories. Today, the heritage scene in Singapore has become vibrant and much effort has been expended to promote awareness of the country's history and heritage. Publications, exhibitions, and CD-ROMs are some of the media adopted to disseminate information.

The NAS has also developed a one-stop Web site called Archives and Artifacts Online Singapore (National Archives of Singapore, 2002) to enable users to browse and search for historical photographs, documents, government speeches, private records, maps and building plans. Users can view videos and listen to the voices of people who have lived through the past. Other activities such as workshops, seminars, open houses, and tours are conducted to attract audiences both young and old.

In addition, the NAS offers reference services to students, researchers and the general public through an archives reference room that serves as a library of research resources. Within the archives reference room, NAS reference officers are stationed at a helpdesk to respond to public enquiries on archival resource matters. Through time, the reference officers noticed that certain enquiries were repeatedly being asked more frequently than others; this usually pertained to well-known subjects or popular topics such as information about the founder of the country, local events during World War II, places of significant historical interest, and so forth. It was then decided that a more systematic approach was needed to cater to such standard enquiries.

The Original Ready Reference System

Physical folders were initially introduced to store related information on particular subjects that were identified by the reference officers as popular enquiries. These were created whenever a reference officer encountered an enquiry on a new subject or topic posed by a user. Hence, each folder would be named or labeled according to the subject based on user enquiries, and information that was compiled to answer that particular enquiry was placed within the respective folder. Related documents and records were filed in chronological order. A simple listing of the document or record titles was sometimes all that was available to help a user identify the individual folder's contents.

The information stored within these folders usually comprised a mixture of actual documents (usually photocopied) and bibliographical records that directed the user to the location where the actual documents were kept. Whenever new information or updates became available, the reference officers had to review the new information, and using mainly his or her judgment, assess whether the new information was related or suitable to be added into existing folders.

To further assist the reference officers in conducting faster searches, a simple Microsoft Access database was installed at the reference helpdesk to complement the existing system of physical folders. The Microsoft Access database was used primarily as a simple enquiry tool to point users to locations of the physical folders. This database attempted to mimic the physical system with one exception – it did not contain any actual documents. It consisted of bibliographical records that had only

three fields – a topic title, an open text field for keying in descriptive data and a remarks field. Much of the information was placed within the open text field and this resulted in miscellaneous information such as location, document identification numbers and other details being lumped together in a non-standardized and haphazard manner. Searching the database became difficult as there were insufficient fields to further distinguish between the records retrieved. In addition, there was no indication from the bibliographical records as to the media type of the actual documents. This is important to some users as they may wish to select only documents of a specific medium.

This rudimentary method worked satisfactorily in the beginning when the amount of information in the database was still small and manageable. However, as the number of different enquiries increased, so did the number of folders, and the reference officers soon found that the Microsoft Access database was unable to support their increasingly complex needs. The database also suffered from one major drawback – it was not available online and users were dependent on an onsite reference officer to access it. This meant that users still had to queue and wait to have their more common enquiries answered. Further, it was also noted by the reference officers that a more systematic way of organizing the information was desirable than the current ad-hoc manner of simply assigning new folders to new enquiries. Subjects and topics could be further be divided and grouped under broader or more specific terms. For example, the main topic of ‘Wars’ can typically be classified as a broad term or subject while more specific topics such as ‘World War 2’, ‘the Surrender of Singapore’ and ‘the Battle of Redhill’ can be cataloged under this as specific terms or subjects.

Requirements Gathering and Analysis

With these issues in mind, a project to develop an enhanced version of the existing database was started. Known simply as the Reference Enquiry Database (RED), the application was conceived to alleviate the current situation faced by the NAS as they seek to improve and automate their reference services as well as to eliminate the necessity of the public to enquire in person at the reference helpdesk for common enquiries.

An important phase of the development of the RED was to assess the existing workflow and identify requirements for the system. During this initial stage of the project, reference officers, who will be the main users of the system, and other stakeholders such as supervisors and management were involved.

Interviews with reference officers were conducted to determine system requirements as well as to ascertain what problems they were experiencing. The interviews gave valuable input to defining and analyzing current and potential problems faced by both the reference officers and members of the public. Some of the significant issues identified were an inadequate and limited Microsoft Access database that reference officers shied away from using, lack of proper organization of information stored in cumbersome files and folders, inability to access the existing database remotely, and difficulty in maintaining/updating the database.

Task analyses were also employed to determine workflow and possible areas of improvement. These included an in-depth study and analysis of functions such as storing and filing of documents, placement and location of documents, searching and

retrieving documents, methods of attending to reference enquiries, identification and classification of common and unusual enquiries, and updating and maintenance procedures.

The analyses found that certain tasks performed by reference officers were relatively inefficient and could be improved or done away with to save on effort and time. For example, even for trivial enquires, reference officers still had to go through the entire routine of formulating a query for the Microsoft Access database, obtain the location of the physical folders, and then examine them to extract the answers. This system could thus be improved by capturing snippets of information that would satisfy general/simple queries within the database records themselves so that there is no need to frequently refer to the physical folders. In addition, folders were also not assigned unique identification numbers during the records creation process and were simply filed according to subject headings, making organizing and searching slow and laborious.

To confirm the accuracy of the information gathered about work patterns and activities of the reference officers, observations were carried out at the archives reference room. The actions of the reference officers on duty, especially when entertaining public enquiries on archival research resources were noted. The observations revealed that the reference officers preferred to depend on their experience rather than use the limited and simple Microsoft Access database that they felt was inadequate. This reinforced the notion that a more sophisticated system was needed to suit their needs.

The Reference Enquiry Database

Users of the RED consist mainly of two groups: the primary group of *internal users*, such as the NAS reference officers who have full access to all of the RED's functions, and the secondary group, the *public*, which is limited only to search and browse functions. The focus here will be on the former group since the procedures for the latter will be identical.

Features

The collation and analysis of the requirements gathered clearly indicated that certain core functionalities, such as the capability to search and browse for folders, add or create new folders and bibliographical records, delete obsolete folders and bibliographical records, and update or modify folders and bibliographical records were a necessity for the internal users (e.g. reference officers) to fulfill their daily reference service duties.

Public users, who are expected to access the database remotely, will also need to have certain features catered for them. The inclusion of instructions, such as how to perform a search with the aid of Boolean operators, i.e. 'AND', 'OR', and 'NOT', should be placed where required throughout the user interface. They serve as an online guide since users will not be able to contact the reference officers for assistance during off-duty hours. In addition, suitable examples and prompts should be placed near unfamiliar input fields. Another useful feature to aid users to quickly sieve through the information available in the database is the ability to browse through the entire collection of folders. Here, it was decided that topics be classified into one or more "specific subjects" which in turn were grouped into "broad subjects".

Data Model

The main purpose of data modeling is to cater to efficient organization and representation of data as well as to formulate meaningful relationships between data. It was decided that the data in the RED be structured in a way that is similar to the current manual system of storing and classifying folders. Specifically, each subject would be represented by a *virtual folder* consisting of one or more *documents*. Documents may either contain actual information that can be immediately used or bibliographic information that point to physical documents.

The metadata structure of the virtual folder and its field names were derived from studies of similar sites such as Collection Finder (Library of Congress, 2002) and the Archival Research Catalog (National Archives and Records Administration, 2002), with the final version customized to suit the NAS' specific needs. Fields used include broad and specific subject to classify the folders, folder description, documents available and their respective media types, and the location of the documents. The latter can either be a physical address for physical documents or a URL for online documents.

The Extensible Markup Language (XML) was adopted for representing the virtual folders in the RED. Among its many advantages are platform neutrality (Yu and Chen, 2001), tag extensibility (Kim and Choi, 2000) and the ability to specify both data and the structure of the data in a format that is both human and machine readable (Gardner, 2001).

Architecture

The RED is a Web-based system and consists of components that are developed around the needs and expected tasks of the intended user groups. These include a Web-based user interface, a module for querying and browsing, modules for folder maintenance, and an XML database server (see Figure 1). All components are developed as Java servlets or Java ServerPages, while the Tamino XML Server (Software AG, 2002) is used to store the RED's records. A Java-enabled Web-server such as Apache is required as well.

Take in Figure 1

The query and browse module is a core component as it not only enables users to query the RED, but is also used by the folder maintenance modules. The module accepts keywords from the user and retrieves and displays relevant folders (and their corresponding documents) matching the query. Boolean and phrase searches are supported.

The folder creation module is invoked when new subjects are identified, requiring new virtual folders to be created. Here, an authorized user is presented with a form containing all the fields in the folder to be completed. JavaScript functions are used to check that all mandatory fields are properly filled before allowing the form to be submitted. The module then generates a new XML document for the virtual folder and stores it into the XML server.

A module for deleting folders is available as well. An authorized user first searches/browses for the desired folder(s) and then selects the folder(s) for deletion. Authorized users may also update existing folders. A desired folder is first retrieved via the query and browse module, and the folder's data is displayed in a HTML form for editing.

To distinguish between public users who can only search the database and internal/authorized users who have access to the RED's folder maintenance features, user validation is required. That is, before folders can be added, updated or deleted, a user must first log into the system by providing a valid user name and password to the user management module. Public users are not required to provide such information in order to search the RED. Finally, to standardize data access and to simplify development, the various modules use a Java-based application programming interface (API) that facilitates communication with the XML server.

CONCLUSION

The main objective of the RED project is to provide the NAS with a Web-based alternative to the current onsite reference helpdesk. The RED is centered on frequently asked reference enquiries from the public that are then matched with archival information to present the user with a convenient online enquiry channel.

Concepts however do not always transform readily into practical solutions and inevitably, problems and issues surfaced that had to be addressed during the development of the RED. One such issue concerns the modularity of the system. In

the current implementation, the RED relies on Tamino for storage and retrieval of its XML records. This has a disadvantage in which a change in the database would require major changes to the existing components of the RED. For example, a version change in Tamino during development had a side-effect that some functions in the data access API no longer worked as expected and had to be written. The problem will undoubtedly be greater, if in the future, the NAS decides to use a database from a different vendor. A better approach is to decouple the data storage mechanism from the rest of system so that the RED can be combined with any database and thus be vendor-neutral. This can be achieved by creating an additional generic data access layer above the current (vendor-specific) Tamino access API so that it shields the RED's other modules from the underlying data storage mechanism.

A second issue is the accessibility of documents. The RED is designed to present both bibliographic data and actual documents of different media types such as text, images, building plans, maps, audio and video. When virtual folders are retrieved, hyperlinks that point to both types of information are provided for users to follow and peruse. However, digitization of documents is an expensive and tedious process so that currently, most of the information stored in the RED is of the bibliographic variety. Consequently, users will still need to physically locate, acquire and browse the actual documents. The realization of the intended level of integration is not expected in the near future.

Nevertheless, the RED offers the potential of not just solving the immediate needs of the reference officers but also to open the NAS' rich collection of archival resources to a global audience via the Web. Remote users can enjoy twenty-four hour online

access to the system and it is expected that existing users will eventually adopt the RED as their primary source of information especially for simple reference enquiries, and utilize the traditional reference helpdesk for more complex information needs.

REFERENCES

- Gardner, J. (2001), "Information architecture planning with XML", *Library Hi Tech*, Vol. 19 No. 3, pp. 231-41.
- Kim, H. and Choi, C. (2000), "XML: How it will be applied to digital library systems", *The Electronic Library*, Vol. 18 No. 3, pp. 183-89.
- Library of Congress. (2002), *Collection Finder*, available at:
<http://memory.loc.gov/ammem/collections/finder.html>
- National Archives and Records Administration. (2002), *Archival Research Catalog*, available at: http://www.archives.gov/research_room/arc/
- National Archives of Singapore. (2002), *Archives and Artifacts Online*, available at:
<http://www.a2o.com.sg>
- Siam, K. (2001), "Transforming centres of excellence – the National Heritage Board experience", *Library Review*, Vol. 50 No. 7/8, pp. 366-73.
- Software AG. (2002), *Tamino XML Server*, available at:
<http://www.softwareag.com/tamino/>
- Yu, S. and Chen, R. (2001), "Developing an XML framework for an electronic document delivery system", *The Electronic Library*, Vol. 19 No. 2, pp. 102-10.

Figures

Figure 1. Architecture of the Reference Enquiry Database