Developing a Digital Library of Reusable Historical Artifacts

Dion Hoe-Lian Goh & Schubert Foo
Division of Information Studies
School of Communication and Information
Nanyang Technological University
Singapore 637718
{ashlgoh, assfoo}@ntu.edu.sg

ABSTRACT

This paper discusses the design and implementation of a digital library of historical artifacts. A major goal of the project is to create an architecture in which artifacts are reusable across various digital library applications. Two such applications have been developed and are described: a virtual exhibition system and a reference helpdesk.

Keywords: Reusable artifacts, digital library applications, archives

INTRODUCTION

The National Archives of Singapore (NAS) is a Singapore government organization that is tasked with the permanent preservation of the country’s records that are of historical importance and significance. These range from private memoirs to government records, and are in various formats including text, photographs, maps, audio and video. To meet the increasing demand for online research, the NAS developed and launched Archives and Artifacts Online Singapore (a²O), a Web portal of heritage information [1]. The portal allows users to search for information across various independent databases and view selected photographs, maps and building plans, listen to oral history interview samples and watch snippets of audio-visual recordings.

A project is underway to extend the a²O portal and develop a digital library of historical multimedia resources offering a greater range of services to both patrons as well as staff of the NAS. In addition to search and retrieval, the digital library will provide basic components for building applications such as virtual exhibitions, lesson development, reference services and annotation tools. The project aims to build a unified platform upon which all of the NAS’ digitized resources may be accessed through various applications, in effect, creating a digital library of reusable artifacts [4]. The eventual size of the collection will be relatively large. There
are for example, currently 1.5 million photographs, 10,000 hours of oral history recordings and 130,000 maps and building plans in the NAS’ holdings.

DIGITAL LIBRARY DESIGN

The digital library is Web-based and consists of two major architectural layers. The *infrastructure layer* consists of components that provide essential digital library services such as artifact and metadata storage and retrieval as well as the repositories for *basic artifacts* – the original resources that form the digital library’s holdings (for example, photographs, maps, video etc.). As different basic artifacts (for example, text versus photographs) may have different requirements for their management and access, this layer supports independent repositories for the artifacts and their metadata. However the digital library offers a single point of access to all artifacts, the infrastructure layer provides a component called the *repository gateway* that serves as a unified front-end to the various repositories. All basic artifact access and management requests will be fulfilled via this gateway.

The *application layer* consists of applications that use the components in the infrastructure layer to deliver services to patrons of the digital library. Applications may range from simple wrappers to infrastructure components such as searching over a repository of photographs, to more complex ones that may combine various infrastructure components to provide value-added services to patrons for example, virtual exhibitions. In many cases, applications may also combine basic artifacts to produce new ones, and these may be atomic (such as annotations) or composite (such as virtual exhibitions). Separate application-specific repositories are maintained for these artifacts, meaning that applications will need to access their own repositories as well as the basic artifact repositories via the repository gateway.

**Implementation**

The infrastructure layer is implemented in Java and consists of application programming interfaces (APIs) for basic artifact access and management. APIs are available for searching, adding, modifying and removing artifacts and their associated metadata. Applications are developed as Java servlets or JavaServer pages and typically include a HTML or Java applet front-end. Metadata and artifacts with textual information are represented as XML documents and stored in XML servers while non-textual artifacts (for example images) are currently stored in the file system.

**APPLICATIONS**

To date, two applications have been developed for the NAS digital library: a virtual exhibition system and a reference helpdesk.

Virtual exhibitions are collections of Web pages revolving around a certain topic. The challenge is to depart from the traditional “copy and use” means of authoring exhibitions to a more efficient “reference and reuse” model. In the former method, artifacts such as text and images
are duplicated when new exhibitions are created. The result is that multiple copies of a single artifact are produced and stored in different locations making updates difficult and storage inefficient. In the “reference and reuse model”, artifacts are stored in a single location (the basic artifact repositories) and referenced in virtual exhibitions [3]. Only one copy of the artifact exists, reducing storage requirements and facilitating changes to the exhibition content simply by changing the artifact(s) involved. In the current implementation, virtual exhibitions are stored as XML documents in an application-specific repository. At execution time, the virtual exhibition system will retrieve the necessary artifacts and convert the XML documents into HTML using an XSL file.

The reference helpdesk provides reference services to common enquiries on Singapore’s culture and history via a Web-based interface. Virtual folders for topics as diverse as life during the Japanese Occupation or water management in Singapore are created and subsequently populated with appropriate references to artifacts through a folder management tool. Users may search the helpdesk or browse the topic classifications until the desired folders are found. These folders are XML documents stored in the reference helpdesk repository and converted to HTML when accessed by the user. However because the digital library is still under development, not all folders will contain artifacts that can be viewed online, although their metadata records may be viewed. Many resources have not yet been digitized and users will still have to make a trip to the NAS to peruse them. Nevertheless, the reference helpdesk is expected to be an improvement over the current situation in which physical topic folders are maintained, requiring users to visit the NAS and speak to one of the staff whenever an enquiry has to be made. In addition, the application will alleviate the heavy workload already faced by NAS staff and allow them to be involved in more complex tasks.

**FUTURE WORK**

Work on the NAS digital library is ongoing with the project focusing on the following:

- **Deployment and testing**: the two applications mentioned previously will be deployed and tested by key NAS staff who will eventually use and maintain these systems. It is expected that through this initial deployment, a better understanding of the usage patterns of the digital library will emerge leading to a further refinement of the digital library’s architecture and its APIs.

- **Metadata for reusable artifacts**: artifacts that can be reused across different applications will likely have different metadata requirements, although a few core fields will remain unchanged. Work is currently underway to evaluate flexible metadata models and approaches for managing them. One approach being considered is through inheritance and/or composition of XML-based schemas.

- **Artifact management**: one near-term project involves an integrated artifact management module for NAS staff to catalog artifacts acquired from government agencies and private collections. Currently, different departments use different systems for cataloging artifacts, leading to potential inconsistencies in assigning metadata as well as interoperability issues across departments.

- **Workspaces and information exchange**: a longer-term project involves the development of workspaces for patrons to maintain personal holdings or to share information within
communities in the digital library. In particular, work will be done on developing “trading protocols” [2] for autonomous exchange of information between patrons’ workspaces.

REFERENCES

1. Archives and Artifacts Online Singapore. Available at http://www.a2o.com.sg/