Management of an AJAX-based Digital Library System

Literature Review

Suraj Subrun

Department of Computer Science

University of Cape Town

sbrsur002@uct.ac.za

ABSTRACT

In order to determine the feasibility of an AJAX-based digital library system, a literature review is conducted regarding the management aspects of such a system. We also explore central repository update features through the use of Web 2.0 technologies. This paper discusses AJAX technologies and provides an overview of the workings of three major digital library systems: Greenstone, Fedora and DSpace. Key concepts and features expected of digital repository management systems are highlighted. Finally, the paper discusses the divide between the Web 2.0 community and the digital libraries community and how Web 2.0 features could be useful applied to the system.

1. INTRODUCTION

The aim of this project is to demonstrate the feasibility of an AJAX-based digital library system. AJAX provides a universal platform for running software. The major advantage of such a system is its portability. This paper is, in particular, concerned with managing and updating the digital library repository.

The following parts of the project are discussed in this article:

1. AJAX technologies;
2. Modifying and administering the information in digital libraries;
3. Updating digital library repositories using Web 2.0 technologies.

Ideally, an AJAX-based system would be able to work straight off a CD-ROM with a standard browser as the only software requirement. To date, the system closest to providing such functionality is the Greenstone Digital Library software suite. Greenstone currently allows users to browse digital library content directly from a CD-ROM but requires a prior installation of several packages before this is possible.

Other important digital library software tools include the Fedora repository management system and DSpace. This paper provides an overview of the methods that these systems use to manage digital libraries.

Finally we have a look at how Web 2.0 technologies can facilitate and promote user created content (UCC) in the field of digital libraries. The key ideas in the concept of Web 2.0 concern interoperability and cooperation and UCC is of major significance (OECD, 2007). Currently, there is a big disconnect between the digital libraries community and the Web 2.0 community (Maslov, Mikeal, & Legett, 2009). The Web 2.0 features that will facilitate contribution to a central repository will also be discussed.

2. AJAX

AJAX (Asynchronous JavaScript and XML) is not so much a technology in itself as it is a group of interrelated technologies in widespread use that are used together to produce highly interactive web applications (Doernhoefer, 2006). AJAX incorporates (Garret, 2005):

- XHTML and CSS for presentation;
- Document Object Model (DOM) for dynamic display and interaction;
- XML and XSLT for data exchange and manipulation;

- XMLHtmlRequest for asynchronous data retrieval;
- and JavaScript, bringing everything together.

The main difference between the classical web interface and AJAX based interfaces is the way data exchange takes place asynchronously (Garret, 2005). In the classical web interface, each time a user chooses a hyperlink, a call to the server is made to request data from the server and this is transmitted back in the form of HTML (Garret, 2005). The user has to wait during the time the request is made and the data is retrieved from the web server. Using AJAX, an engine is loaded initially and this engine can simultaneously render an interface for the user and communicate with the server (Garret, 2005). Therefore, whenever new data has to be retrieved, this does not stall user interaction (Garret, 2005). The data is normally transferred in the form of XML from XML or HTML servers (Garret, 2005).

One major limitation of the JavaScript language is its lack of file access functions. The main reason for the omission of this feature is to prevent security violations on the client's machine (Flanagan, 2002).

3. EXISTING DIGITAL LIBRARY SYSTEMS

One of the leading digital libraries systems is Greenstone (New Zealand Digital Library Project, 2009). Greenstone is a tool for building digital libraries that "provides a new way of organizing information and publishing it on the Internet in the form of a fully-searchable, metadata-driven digital library" (New Zealand Digital Library Project, 2009). Greenstone requires several packages to work, for example scripting and web server packages (New Zealand Digital Library Project, 2009).

Greenstone works by ingesting metadata (supporting various standards such as Dublin Core, RFC 1807, etc) and various types of digital resources, using different plug-ins for various document formats, to produce its own set of XML data files (New Zealand Digital Library Project, 2009). These are subsequently converted into searchable indexes (New Zealand Digital Library Project, 2009). As mentioned earlier, Greenstone provides allows a digital library to be compiled onto CD-ROMs that can be distributed and run autonomously (New Zealand Digital Library Project, 2009). These CD-ROMs however require the installation of several software packages before being used (New Zealand Digital Library Project, 2009).

The Fedora Project is a digital object repository management system. Its architecture is based on object models that are templates data objects, the units of content (digital resources and metadata) (Staples, Wayland, & Payette, 2003). Behavior objects are used to describe the operations of tools and services on the data. Behavior objects are themselves described in the form of metadata (Staples, Wayland, & Payette, 2003).

The Fedora repository system consists of three layers: the Web Services Exposure Layer, the Core Subsystem Layer, and the Storage Layer (Staples, Wayland, & Payette, 2003). The Web Services Exposure Layer provides separate interfaces for management and access of the digital objects in the repository (Staples, Wayland, & Payette, 2003). The Core Subsystem Layer is responsible for the management and access subsystems (Staples, Wayland, & Payette, 2003). The management subsystem implements the operations needed for creating, modifying, deleting, importing, exporting, and maintaining digital objects (Staples, Wayland, & Payette, 2003). It also caters for validation and integrity of data (Staples, Wayland, & Payette, 2003). The access subsystem provides the methods for showing the content of digital objects (Staples, Wayland, & Payette, 2003). The Core Subsystem Layer also consists

of a security subsystem with policy management and enforcement (Staples, Wayland, & Payette, 2003). The last layer, the storage subsystem, deals with reading, writing and removal of data from the repository. Digital objects are stored as XML-encoded files (conforming to METS schema) (Staples, Wayland, & Payette, 2003). The Fedora repository system also allows flexible relations between digital objects to be stored and queried (Fedora Development Team, 2005).

DSpace, another digital library solution, is made up of communities which contain groupings of related content defined as collections (The DSpace Foundation, 2009). These collections consist of items which are the basic archival elements (each using Dublin core metadata record for identification) of this system (The DSpace Foundation, 2009). Items may appear in different collections but have only one owning collection (The DSpace Foundation, 2009). An interesting aspect of DSpace is that each item has a bitstream containing data defining the digital object but, in addition, has an associated bitstream format (The DSpace Foundation, 2009). This is intended to cater for the preservation of data. Bitstream format data is more specific than MIME types or file suffixes as, for example, *application/ms-word* or *.doc* files can contain different formats depending on the version of Microsoft Word the file was created with (The DSpace Foundation, 2009). Furthermore, each bitstream format has a support level to indicate how well the hosting institution is likely to preserve content in the format in the future (The DSpace Foundation, 2009). Other features of DSpace include access control to certain of its features (The DSpace Foundation, 2009).

After considering the workings of these three systems, we observe that the main trend is to provide an abstraction to digital objects (such as "items" or "data objects") and affix them with metadata before grouping and organizing them into searchable collections. Features which are required of digital library software include accepting various standards of metadata, such as Dublin Core, providing security features (access control), and making data accessible and searchable through web interfaces. Storage layers usually wrap the data in custom XML files. Optional features include storing information about relations between digital objects, storing file formats in the hope of long-term preservation and independent CD-ROM distribution of the digital libraries.

4. WEB 2.0

Web 2.0 is a broad term and is often considered to be just a "buzzword". Its validity is still being argued since most of the technologies that make it up are not new. The general tendency however is to consider Web 2.0 as a useful abstraction referring to a guiding set of trends and practices towards producing better applications (Maslov, Mikeal, & Legett, 2009).

Applying the Web 2.0 concepts to digital libraries often results in a conflict between cooperation and control (Maslov, Mikeal, & Legett, 2009). On one hand, Web 2.0 practices that allow external participants to contribute information engender several problems, for example establishing ownership of information. On the other hand using the more conservative approach, content is easily maintained but is closed to the rich external pool of information (Coombs, 2007).

Web 2.0 provides several features that enhance users' experience. AJAX technologies provide rich user interfaces that mimic familiar desktop applications (Coombs, 2007). Remixable content is another key feature of Web 2.0 whereby application programming interfaces are provided to allow the content to be meshed with content on other websites (Coombs, 2007). Another key concept of Web 2.0 is the use of

lightweight programming models (O'Reilly, 2007). This deals with the way Web 2.0 applications use data, focusing on producing useful output rather than the processing of the data (O'Reilly, 2007).

The advantages of Web 2.0 include a richer user experience for all the users of the system because they can participate and they can experience a larger variety of content on the site (Coombs, 2007). Moreover, AJAX provides users with a more desktop-style-application experience for the users (Coombs, 2007).

The interactivity of Web 2.0 features will be beneficial to digital library systems by providing the technology to communicate data to central repositories. The way in which Web 2.0 promotes UCC can be seen as a step towards richer and more engaging digital library content.

5. SUMMARY

We have discussed the advantages of AJAX technology and its potential application in creating a lightweight digital library system. The various technologies that make up AJAX are widely available and successfully creating such a system would have the important benefits of portability and preservability.

The research into the various existing digital library solutions has provided us with the key features that are expected of a modern digital library system and will have to be implemented in our project to demonstrate the effectiveness of an AJAX-based solution.

We have also seen that Web 2.0 provides the technology for communications but in addition to that has the ability to encourage richer contributions and library data.

REFERENCES

Coombs, K. A. (2007). Building a Library Web Site on the Pillars of Web 2.0. *Computers in Libraries , 27* (1).

Doernhoefer, M. (2006, July). Surfing the Net for Software Engineering Notes - JavaScript. *ACM SIGSOFT Software Engineering Notes , 31* (4), pp. 16-24.

Fedora Development Team. (October 28, 2005). *Fedora Open Source Repository Software, White Paper.*

Flanagan, D. (2002). *JavaScript: the definitive guide* (4th Edition ed.). O'Reilly.

Garret, J. J. (2005, February 18). *Ajax: A New Approach to Web Applications*. Retrieved May 2009, from Adaptive Path: http://adaptivepath.com/ideas/essays/archives/000385.php

Maslov, A., Mikeal, A., & Legett, J. (2009). Cooperation or Control? Web 2.0 and the Digital Library. *Journal of Digital Information , 10.*

New Zealand Digital Library Project, University of Waikato. (2009). *Greenstone Factsheet*. Retrieved May 2009, from Greenstone Digital Library Software: http://www.greenstone.org/factsheet

O'Reilly, T. (2007). What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications and Strategies* (65), 17-36.

Organisation for Economic Co-operation and Development (OECD). (2007). *Participative Web and User-created Content - Web 2.0, Wikis and Social Networking.* OECD Publishing.

Staples, T., Wayland, R., & Payette, S. (2003). The Fedora Project, An Open-source Digital Object Repository System. *D-Lib Magazine , 9* (4).

The DSpace Foundation. (2009). *DSpace Manual.* Retrieved May 2009, from DSpace Website: http://www.dspace.org/1_5_2Documentation/index.html