

Research and Implementation of Content-management Based on FEDORA

Yongchang Wang, Ligu Zhu

School of Computer Science, Communication University of China, Beijing, 100024, China
Shijiazhuang University, Shijiazhuang, 050035, China

Keywords: non-structured data, FEDORA, digital object, content-management.

Abstract. The surge of information leads the creation of non-structured data increasingly, so content-management has been one of hot topics for IT industry to research. This paper discusses the use of FEDORA in content-management by analyzing the digital object model in FEDORA. We implement a content-management system based on FEDORA. This system combines some real digital object that can be used in FEDORA and tests its performance.

1 Introduction

When we mention the problem of data management [1], we always think of database system. However, database can only deal with structured data that take 15% of all data, so how can we manage non-structured data of 85%, that is responsibility of content-management system.

Information explosion gave birth to content management in today's era of information explosion, information is growing at an alarming rate everyday. Statistics show that, data from the transaction in the annual growth rate reached 61%, and the annual growth rate of other information related even exceeded 92%. At the same time, the types of information are also extended constantly, unstructured information appear constantly more and more, including various enterprise statements, bills, electronic documents, pictures, site of various elements, fax, scanning images, and audio, video, and other multimedia information etc. But behind the information explosion, the challenge we have to face is how to orderly storage, management and use of these information, especially for those large amounts of unstructured information. If an enterprise can not put these information systematically cataloging and management, it will face the crisis of constantly losing productivity. Another statistics is worth us thinking deeply, if an enterprise staff want to complete his work, the time to search for information will cost 30% of the whole work time, this will reduce the employees' work time from 40 hours to 28 hours a week. No company allows employees to leave with 30% time annually, but now, this is the fact that a lot of enterprises and organizations have to face with. On the other hand, enterprises and organizations want to have more control of the information, ensuring information security, mining the value of information.

The content is also known as the digital content, are generally defined as electronic information that anyone will use in business and personal lives everyday. It can be all kinds of digital information, including documents, publications, image, audio, video, software, forms, receipts, bills, design, contract, and e-mail. Content management technology will store different types of digital content properly, and make full use of adequate information to query and retrieve digital assets stored efficiently. Data mining technology and the intelligent analysis make the digital content can be fully utilization, the value of product will continue to improve, this will enhance the competitiveness of enterprises in the information age.

Many enterprises or organizations want to take more control of these non-structured data, to ensure the data's security, and to mine the value of these data [2]. After years of effort, many technologies of content management has been made much progress, and there are lots of content-management software has appeared so far. However, there are still many problems to be resolved, they are as follows [3] [4]:

- These content-management software mainly focus on some specific type of resources. based on such platform, we can only develop some specific application systems such as image database or a document management system.

- They can not represent the relation of resources effectively. A common problem in many content management systems is the lack of an expression of relationship between resources effectively, some systems even ignore the relationship between resources, or even regard the resource as a solitary, similar to the document object.
- It is hard to implement the cooperation among systems due to the lack of standard interface.
- They are lack of desired flexibility and scalability. At the initial stage of implementing the content management, system developers often rarely understand the requirements of their customer, and the customer's demand is always changing dynamically. If the design of systems is based only on these specific requirements, it is difficult for such systems to adapt to the future changes.
- The use-cost for users is still very high.

The main reason that cause the problems above lies in the data object what they are using. If we want to solve these problems, we should focus on the data object that can be used in content management. Therefore, we can use a free software called FEDORA as the engine of content-management.

2 Data object and its operation mechanism based on FEDORA

2.1 Introduction to FEDORA

FEDORA (Flexible Extensible Digital Object and Repository Architecture) is originally a part of research project in Cornell University [5]. It was used in digital library as an open frame. In the year of 2001, Virginia University and Cornell University jointly redesigned FEDORA with support of the Mellon Foundation. From then on, FEDORA has a series of version quickly, and the latest version is FEDORA 4.X[6].

2.2 The basic characteristics of FEDORA

- Open source software, download address is: <http://www.fedora.info/download/>.
- Written by the Java, supporting the popular operating system such as Windows, Linux, Unix etc.
- Extensive database support, FEDORA has not only comes with built-in database McKoi, and now support popular databases ,including MySQL, Oracle8i, PostgreSQL, Oracle9i etc.

2.3 Data object model based on FEDORA

FEDORA has two basic bodies [7]: Digital object and Repository. Digital object is the kernel of the entire frame and it is the key to understand FEDORA frame. Repository provide service for digital object, it provide the function of deposition, copy and storage for digital object. Digital object model of FEDORA is widely concerned about by the content management field with its extensibility, flexibility and reusability.

Generally speaking, a digital object is a digital resource that has a unique Identifier. It consists of Persistent Identifier, Disseminator, Data stream and System Metadata. Each digital object has only one persistent identifier. Each persistent identifier in the FEDORA repository will uniquely identifies a digital resources, and provides the technical guarantee to work in parallel for multiple resource library. Each digital object can have one or more data streams and distributors. Every digital object must have system metadata, because the repository will use it to operate digital object.

The most significant characteristic of FEDORA digital objects are: digital object is polymerization of the content (including data and metadata) and behavior (such as services), and both are available through a uniform resource identifier URI (Uniform Resource Identifier) to release

Figure 1 shows the structure of a digital object model [8] [9].

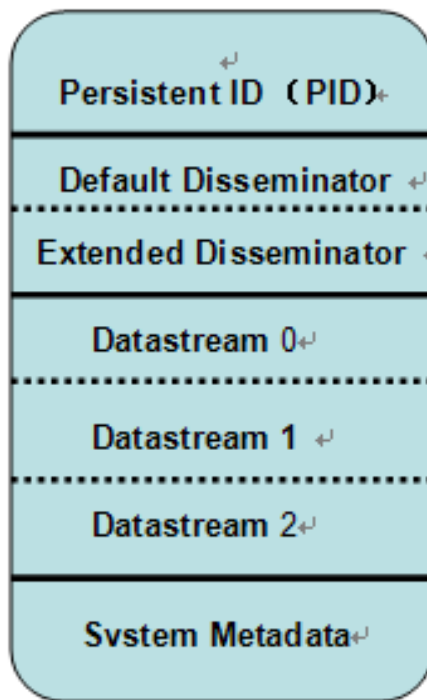


Fig.1: digital object model

2.4 Operation mechanism of a digital object model

Content type distributor digital object model required for the operation of data objects, it includes three kinds of information: The directional information of description of service operations (i.e. connected on the behavior of the object is defined in the service description); directional information of description of service (i.e. connected on the behavior mechanism of object identifiers); related data stream identifier.

The procedure of processing a digital object model is[10]: When a client request a disseminator of a digital object model ,the disseminator get the information which service point to and related data stream ID, read the binded relation between service and data stream , then read the data stream and send it to server for processing, at the result will return to client at the end. This procedure can be described as figure 2.

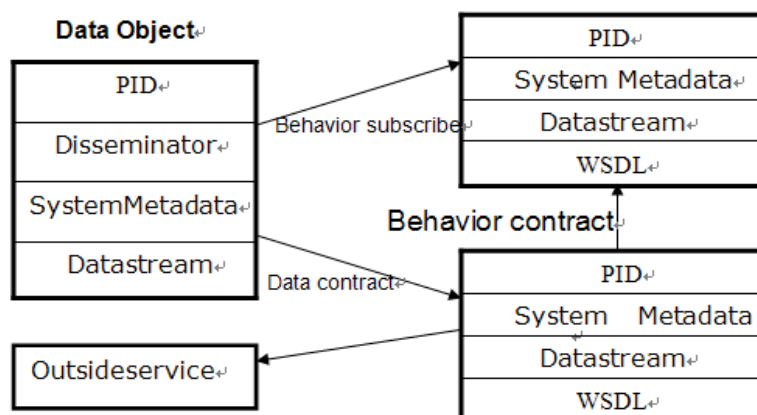


Fig.2: operation mechanism of a digital object model

3 The design and implementation of a content management system based on FEDORA

FEDORA is written by JAVA, so our content management system can use the J2EE architecture naturally. Our system is based on B/S structure combined C/S. The web service of the system can

be built-in Tomcat or some popular web service such as IBM WebSphere or BEA WebLogic. The whole system can be divided into three layers: client layer , business layer, and DB layer, which is showed by figure 3

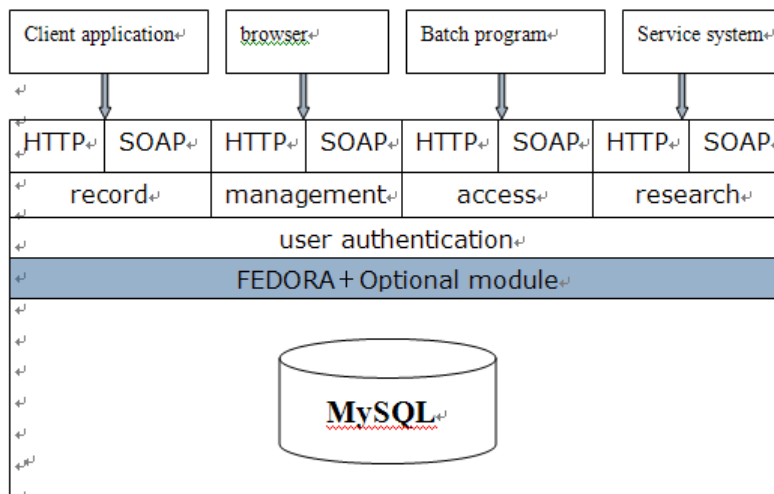


Fig.3: system structure

Client layer: The C/S interface that is written by JAVA SWING serves as a tool for recording the metadata and the disseminator of a digital object, as is shown by figure4. While the B/S interface that is written by Jsp serves as a interface for searching and accessing digital resources, as is shown by figure 5.

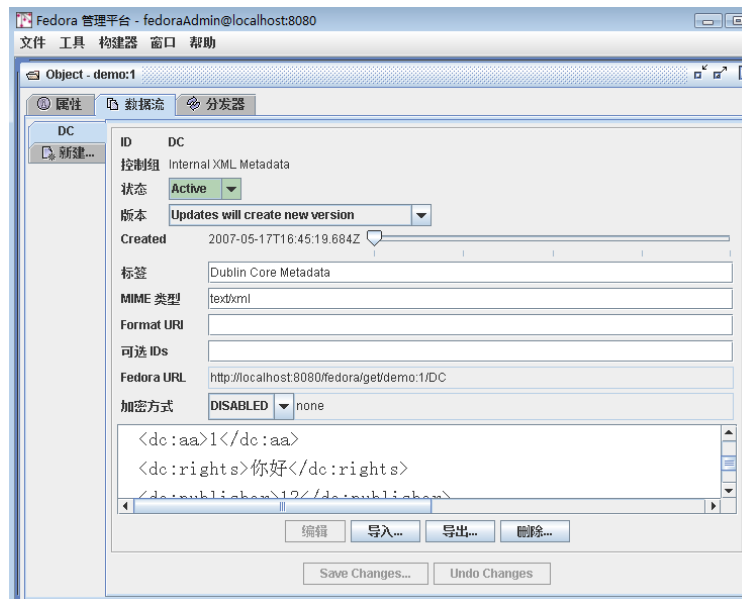


Fig.4:interface based on C/S



Fedora Repository

Find Objects

Fields to display:

<input checked="" type="checkbox"/> pid	<input type="checkbox"/> bDef	<input type="checkbox"/> type
<input type="checkbox"/> label	<input type="checkbox"/> bMech	<input type="checkbox"/> format
<input type="checkbox"/> fType	<input checked="" type="checkbox"/> title	<input type="checkbox"/> identifier
<input type="checkbox"/> cModel	<input type="checkbox"/> creator	<input type="checkbox"/> source
<input type="checkbox"/> state	<input type="checkbox"/> subject	<input type="checkbox"/> language
<input type="checkbox"/> ownerId	<input type="checkbox"/> description	<input type="checkbox"/> relation
<input type="checkbox"/> cDate	<input type="checkbox"/> publisher	<input type="checkbox"/> coverage
<input type="checkbox"/> mDate	<input type="checkbox"/> contributor	<input type="checkbox"/> rights
<input type="checkbox"/> dcmDate	<input type="checkbox"/> date	<input type="checkbox"/> aa

Search all fields for phrase: [help](#)

Or search specific field(s): [help](#)

Maximum Results: 20

pid	title
demo.1	Enter a one-line description of the object.
changeme.1	Enter a one-line description of the object.

Fig.5: interface based on B/S

Business layer: This layer can be composed of FEDORA and other optional module, such as workflow engine. FEDORA is responsible for deposition, copy and storage of a digital object. It also provide an API for client and interact with DB.

DB layer: Although FEDORA has a built-in database called McKoi, we still choose Mysql as our database for storing some properties of digital resource, because we have to considering the scalability of our system.

4 System performance analysis

FEDORA support the function of batch import of digital objects. We made a test for system performance. In the test, we have totally imported 20000 digital objects, including the whole import and its copy import. Figure 6 shows the change of time consumption per 50 digital objects. We can infer from figure 6 that time consumption is in proportion to the number of digital objects.

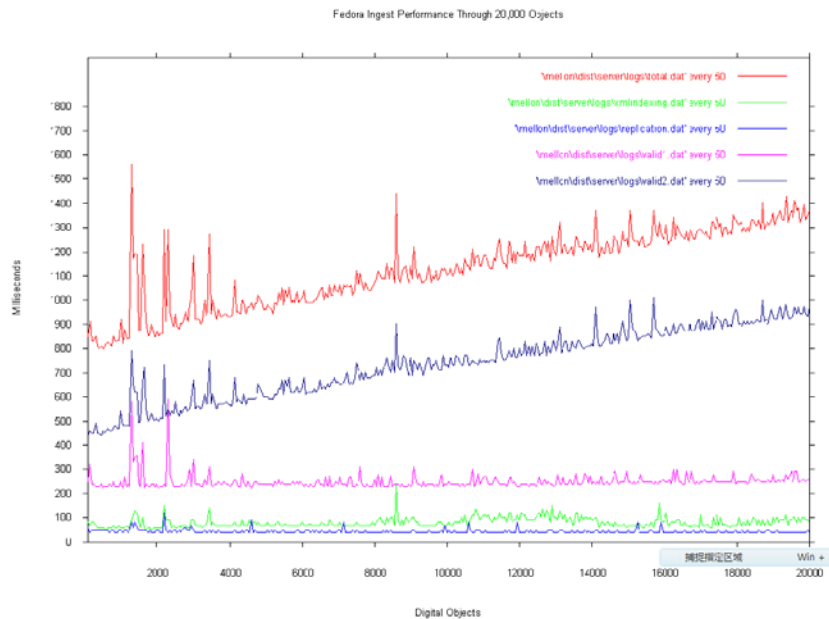


Fig.6: Performance test diagram

5 Conclusion

FEDORA has a key feature of scalability and reusability. It has been widely used in content - management of website, digital library, file storage system, and so on. At the same time, FEDORA has valuable reference to other systems.

6 Acknowledgements

In this paper, the research was sponsored by “The Comprehensive Reform Project of Computer Science (Project No. ZL140103)” and partly supported by “The Engineering Planning Project of Communication University of China (Project No. XNG1436)” .

References

- [1] Digital library research : current development s and t rends ,Ali Shiri ,2003
- [2] A Framework for Dist ributed Digital Object Services ,Robert Kahn ,Robert Wilensky , May 13 , 1995
- [3] Key Concept s in the Architecture of the Digital Library ,William Y. Arms , J uly 1995
- [4] Implementation Issues in an Open Architecture Frame2 work for Digital Object Services ,Carl Lagoze ,David Ely ,J une 6 ,1995
- [5] An Architecture for Information in Digital Libraries ,Wil2 liam Y. Arms , Christophe Blanchi , Edward A. Overly ,February 1997
- [6] A Metadata Architecture for Digital Libraries , Ron Daniel J r. ,Carl Lagoze ,Sandra D. Payette ,1997
- [7] Dist ributed Active Relationships in the Warwick Frame2work , Ron Daniel J r. , Carl Lagoze ,1997
- [8] Interoperability for Digital Object s and Repositories : The Cornell/ CNRI Experiment s , Sandra Payette , Christophe Blanchi ,Carl Lagoze , Edward A. Overly , May 1999
- [9] Virginia Dons FEDORA :A Prototype for a Digital Object Repository , Thornton Staples , Ross Wayland , August 2001
- [10] The Mellon Fedora Project :Digital Library Architecture Meet s XML and Web Services , Sandra Payette and Thornton Staples ,2002