Research on Integration Scheme and Framework of Public Digital Cultural Resources

Chao Wang

Information and Educational Technology Center, Southwest Minzu University, Chengdu, Sichuan, China

Abstract. The integration of public digital cultural resources refers to the integration of digital resources distributed in museums, libraries, archives and other public service institutions, to achieve one-stop resource search and acquisition through clustering, integration and reorganization. This study mainly discusses the main problem facing the public digital cultural resources integration system is how to build an open and interconnected three-tier digital resources integration scheme based on the digital resources owned by multiple heterogeneous public cultural research platforms through interoperability and data sharing between platforms, so as to realize cross-platform, cross-language and the "one-stop" access mode of users.

1 Introduction

Culture is all the result of human creation. Public digital culture is a kind of culture produced under the digital environment, which belongs to the category of public culture and it is a combination of public culture and digital culture as well [1]. Today, with the rapid development of information technology, public cultural service institutions including museums, libraries, archives etc. have built a number of information service platforms to match their own application needs based on their respective business requirements, consumers, functions and other considerations. Managers hope to upgrade the level of digital construction through the use of the platform. The integration of public digital cultural resources refers to the integration of digital resources distributed in museums, libraries, archives and other public service institutions, through clustering, integration and reorganization, to achieve one-stop resource search and acquisition [2]. In the implementation of integration projects, several questions are raised in digital resource interaction on heterogeneous platforms, standardization of resource processing and diversity of user needs.

2 The design of scheme and framework

2.1 The introduction of XML

XML is short for eXtensible Markup Language which is a syntax requirement proposed by W3C (World Wide Web Consortium) to standardize and unify the organizational format of information and facilitate the interaction of data between applications in the network. XML as a meta-markup language, users can freely and

flexibly define the markup language, element content, data type and so on according to their actual needs under the premise of meeting the syntax rules, so as to accurately express the specific meaning of information contained in each markup in the process of events. In addition, XML as a way of representing data, the whole document structure adopts tree pattern, and uses its format to store various information. Each node of the tree structure is each object in the event, and the object can be mapped to the corresponding element, which is a collection of tags and content. As a semi-structured language, XML describes information through nested relationships between elements. Because of its tree-like and hierarchical text structure, XML always maintains this structural relationship in the process of data transmission. When multiple applications share or parse the same XML document, there is no need to use traditional string parsing or disassembly process [3]. Compared with HTML (Hypertext Markup Language), although HTML is widely used, it has long been difficult to rely solely on one of its file types to face the changing data content. HTML also uses a platform-independent text file format in which text content is also composed of tags, syntax is simple and easy to use with powerful functions. However, the initial purpose to designed HTML is to display data content and structure styles, primarily to facilitate user experience and browsing. XML, on the other hand, separates data from display styles, focusing on describing data content to multiple applications and storing data content through text tags.

Generally speaking, as a technical support, XML has its own functional characteristics to achieve the integration of public digital cultural resources provides the possibility. Firstly, XML users can set up markup language independently in a specific domain, and use

Schema, DTD and other verification mechanisms to standardize the format of elements and attributes, so as to avoid semantic conflicts as far as possible, to ensure the interaction of information. Secondly, XML can be written in pure ASCII text. The content of the data and display format are completely separated. Users can adjust the data itself or the browsing interface separately. At the same time, because of the ASCII text format, some missing XML storage content will not completely affect the data parsing. Finally, XML can easily convert object files of different formats through XSLT language, and has the ability to run on various heterogeneous system platforms and application software.

2.2 The design of the framework

The purpose of the public digital cultural resources integration project is to provide individualized and differentiated information organization to the public based on the digital resources owned by many heterogeneous cultural institutions through the establishment of the public digital cultural resources integration platform. same time, to use semantic tools of information organization to carry out its resources. In the meantime, it can also make use of information organization semantic tool to process its resources and standardize them [4]. So as to achieve the "one-stop" lookup and acquisition of digital resources by users. The whole system is divided into three layers, - the basic data layer, the application service layer, and the platform presentation layer. It is shown in the figure 1.

Figure 1. Integration Framework of Public Digital Cultural Resources.

The basic data layer contains digital resources owned by multiple cultural research institutions, which is the basis of the entire resource integration platform. Cultural research institutes such as libraries, archives, museums and research institutes keep various types and formats of data according to their own characteristics, such as the fields of industry, functional attributes and business categories [5]. Generally speaking, there are three types: structured data, unstructured data and semi-structured data. Structured data, also known as row data, can be logically expressed and implemented by a two-dimensional table structure. Data formats and

specifications must be strictly followed. Generally, relational databases are used to store and display data, such as library catalog search, archives storage, and financial statements. This kind of data can be easily accessed and queried by users because of its "rules" to follow. For its standard, the data exchange between platform and platform is easy to implement. The data model of unstructured data cannot be defined in advance, and cannot be described by two-dimensional logical table of database, such as text, picture, audio, video, report forms, etc. These data are generated from communications, media, websites, sensors and so on. In the information age, these applications are widely used, because they are easy to store, to express. Public cultural service organizations also store related types of data, such as electronic books, archives of literature and history, museums, cultural relics, pictures, research institutes, and so on. However, due to the randomness and uncontrollability of such data, it is difficult to implement data analysis, data mining and data interaction between heterogeneous platforms. Semi-structured data is mainly used to identify the elements of the internal markup, to achieve the hierarchical division of information, convenient for node storage, such as EDI, NoSQL, XML, HTML and so on. With the continuous development of information technology and the gradual application of artificial intelligence technology, the proportion of semi-structured data resources in public digital cultural service organizations will gradually increase.

The application service layer, which is located between the basic data layer and the platform presentation layer, is the core of the public digital cultural resources integration system. The main function of the application service is to deal with heterogeneous data in different cultural institutions within the system. The Application module includes the receiving of user's request, data acquisition, analysis, encapsulation, transmission, request feedback and so on [6]. The process is based on user requests submitted on the interactive platform. Through the wrapper, the basic data source data is collected, the heterogeneous data is generated into standard XML document according to the unified global pattern, and then the data in the XML document is extracted by SAX/DOM parser. Finally, the final data is returned to the public resource platform through filtering, cleaning, clustering, fusion and reorganization. In the application service layer, the function of wrapper, SAX (simple API for XML, SAX) and DOM (Document Object Model, DOM) parsers is especially important. The wrapper uses Web Service technology. Web Service is a standardized, platform-independent, low-coupling, programmable application oriented to the Internet. The application can also be described, configured and published in XML language. It is mainly composed of three components: SOAP, WSDL and UDDI. Using the interaction of these three components, users' requests can be transmitted to heterogeneous data source system. Then the retrieval results can be returned to the parser by searching the database of the system. That is, the requests and retrieval results of the data can be sent and uploaded. Data is passed to the parser, which is used to retrieve and configure elements, attributes, and content in an XML

file. The parser is generally composed of SAX or DOM. SAX adopts the event-driven model. It does not need to traverse the entire XML document when parsing the XML document. When parsing to a given label, it automatically activates the callback mechanism and the parsing process ends. The advantage of SAX is that parsing can be terminated at any time according to label, so processing large documents can improve efficiency. But the disadvantage is that it cannot access many data in the same document at the same time, and the encoding process is more complicated. The DOM parser transforms an XML document into a tree containing its contents, each node of which is the data itself. The parsing process reads the XML document of the logical tree structure into memory, and then traverses the entire document. DOM has the advantage of random access to any element or content, and supports modification of content. The drawback is that as the content of the document increases, the memory footprint increases and the operation efficiency decreases. Therefore, when choosing a DOM or SAX parser, users don't have to stick to a pattern, but they can choose according to the specific situation.

The platform presentation layer is located at the top of the integration framework, which mainly provides users with friendly interactive operation interface. The platform adopts B/S or C/S interactive mode. By accessing the public digital cultural resources integration platform, users can automatically collect and summarize digital resources of heterogeneous organizations that have been integrated into the platform by searching a key word, sentence, or topic as a clue. Content is fed back to users by information organization such as directory list, spatial model and visualization tools [7]. Public digital cultural resources integration platform collects digital resources owned by public cultural service institutions such as libraries, archives, museums, research institutes, and so on, and processes them through interactive, cleaning, screening, mining and other information technologies [8]. According to the needs of the previous research, and combined with the actual application at home and abroad, the author thinks that the integration platform function module should include the following aspects of service content: (1) Digital resource navigation system. Users can customize any search path in digital resources according to the characteristics of the information to be searched, and accurately find the task target. It can also be pushed through platform and column display to facilitate data acquisition by users. (2) Multi-class retrieval system. In view of the difference of user groups, the retrieval system can be divided into three types: simple, complex and multi-lingual. That is to say, we should not only consider the simplicity and ease of use of the daily needs of ordinary people, but also take into account the accuracy and strictness of scientific research by experts and scholars, and also the accuracy and efficiency of multi-lingual and cross-lingual translation. (3) Unified information portal. To construct a unified information portal website, users can access a single entrance to the website through Web browsers. The platform uses XML technology to eliminate logical or physical isolation between heterogeneous systems, enhance human-computer interaction, open resource sharing, provide

personalized and differentiated services for different occupational groups, and truly realize the "one-stop" user experience.

3 Model

One of the first problems facing the integration of public digital cultural resources is that there are many industries, departments, or research institutes, which have different data formats, data standards and types of databases among various information systems and business platforms. How to shield the heterogeneous data sources and establish a unified global model under the premise that the underlying data environment is very different, so as to realize the seamless connection between the business system and the integration platform, and the safe and fast free interaction of data. In this paper, XML Schema function is used to establish a model driven mapping method.

The XML Schema is a text file with an extension of ".xsd" and can be written in the XML language. Its function can not only describe the structure and content of documents, define the relationship between elements and elements, but also constrain the data types of elements and attributes in text. Using Schema technology, while establishing mapping relationships, the ability of defining data types using Schema can make it easier and more detailed to describe text content, verify the accuracy of data, and transform between different data types. Taking the bibliographic retrieval as an example, through the establishment of the previous bibliographic model, each field in the model is matched with each field in the bibliographic list in the database to be retrieved, and the mapping relationship is established. Then the field in the bibliographic model is replaced by the bibliographic form field in the original database, and the original data content is assigned to the model field. Finally, XML language is used to describe it in a unified way, so as to eliminate the problems of data type difference and semantic inaccuracy between heterogeneous platforms. Part of the contents are as follows:

Bibliographic form in original database, Table001
A bibliographic model, Pattern001

Title	Author	Press	Data	Price
My Memories of Old Beijing	Haiyin Lin	The Commercial Press	2017-10-01	11.80

A bibliographic model, Pattern001

<Test>

SM	ZZ	CBS	RQ	JG
----	----	-----	----	----

```
<Pattern name="Pattern001">
  <Table name="Table001">
    <SM>Title</SM>
    <ZZ>Author</ZZ>
    <CBS>Press</CBS>
    <RQ>Data</RQ>
    <JG>Price</JG>
  </Table>
</Pattern>
```

</Test>

<Pattern001>

<Bibliography>

<SM>My Memories of Old Beijing</SM>

<ZZ>Haiyin Lin</ZZ>

<CBS>The Commercial press</CBS>

<RQ>2017-10-01</RQ>

<JG>11.80</JG>

</Bibliography>

</Pattern001>

After the above XML language expression, Title, Author, Press, Data, Price and other fields in the original bibliographic form are corresponding and assigned to the title, author, publisher, date and price of the book in the model, and the model field is replaced by the form field. The data after association not only clarifies the data type, but also reduces the semantic ambiguity, thus ultimately guaranteeing the consistency of user understanding. For example, 2017-10-01, because it constrains the "date" of the data type in XML format requirements for "YYYY-MM-DD", through this definition, has accurately described the true meaning of the data content.

4 Summary

The construction of public digital cultural resources is an important part of China's public cultural system. The emergence of a large number of heterogeneous data is inevitable in the construction of resources. This paper uses XML technology, through the framework design and model architecture of the integration platform, hoping to be able to realize the public digital cultural resources integration system cross-platform, cross-language, and data independent interaction of the "one-stop" user access service mode.

Acknowledgement

This research was financially supported by the Key Research Base of Philosophy and Social Sciences in Sichuan: Qing Zang Plateau Humanities and Social Sciences ("Research on Integration Scheme and Framework of Public Digital Cultural Resources", Grant NO. QZZ1805).

References

1. Miao Wang, Honglei Sun, Jianmin Zheng, Public Digital Culture: an Analysis of the Concept and the Research Development, *Journal of Modern Information*, vol.37, pp.172-177, 2017.
2. Mintang Hu, Jianming Zheng, Research on the Connotation, Current Situation and System of Public Digital Culture Construction, *Documentation, Information & Knowledge*, vol.06, pp.32-38, 2012.
3. Zhongyu Tian, Lei Guo, XML Practice Course, Tsinghua University Press, 2016.
4. Chunzi Yan, Construction of Public Cultural Digital Resource Aggregation Service Platform, *Researches in Library Science*, vol.11, pp.45-47, 2016.
5. Yi Tang, Ximing Xiao, Lihong Zhou, On the Model Building of Public Digital Cultural Resources Integration in China, *Library Journal*, vol.07, pp.12-25, 2016.
6. Rong Tian, Yi Tang, Resource Organization Mode in the Integration of Foreign Public Digital Cultural Resources, *Information and Documentation Services*, vol.06, pp.68-73, 2016.
7. Xiaohe Ma, Sha Li, Application of XML Technology in Big Data Environment, *Wireless Internet Technology*, vol.02, pp.112-114, 2018.
8. Hongbin Tang, Junsong Liao, Sheng Di, Research on Design of E-commerce Application System on XML, *Journal of Computer Applications*, vol.22, pp.15-21, 2002.