

# Collection and storage of digital web files based on intelligent management information system

Dong Wang, Wu Zeng

China Academic of Electronics and Information Technology, Beijing, China

**Keywords:** information construction; information management; Web; university library

**Abstract:** Based on the new information management concept of information construction, this paper analyzes the process management of modern university library from two aspects of logistics and information flow, and puts forward the overall design scheme of Web information management system including five main modules: Web site, public retrieval system (OPAC), web information management system, virtual consulting service system and information feedback evaluation mechanism. The purpose of this paper is to build a clear and usable web information space for users. The intelligent information collection and processing system based on web is studied in this paper. On the one hand, it adopts efficient URL de duplication and template based download mechanism, which greatly improves the performance of collecting web resources; on the other hand, it applies mature and advanced natural language processing technology to intelligently classify and summarize the collected information.

## 1. Introduction

Today's information users use the library for information query, there are two entrances: the traditional building entity entrance and the virtual web entrance. The traditional building entrance refers to the information service received by users who go to the specific library to borrow and inquire in person, while the virtual web entrance refers to the online information service that users can enjoy when they log in to the library network web entrance by using the computer network. For today's library managers, providing a reasonable and easy-to-use and humanized information space for information users is the key to achieve organizational goals. Therefore, on the one hand, the managers of the organization need to create a good library building environment, academic environment and excellent organizational culture, on the other hand, they need to solve how to process, arrange, effectively organize and display the existing information resources of the library and pass them on to users. Based on this, this paper attempts to design the web information management system of University Library Based on the team concept, so as to build a clear and usable web information space for users and create a good web information ecological environment.

## 2. System architecture

The system consists of three subsystems: Web information collection subsystem, information intelligent processing subsystem and information publishing subsystem. The three subsystems can be deployed and run separately, and can also realize the automatic collection, intelligent processing and active release of the whole process through the interface file.

### 2.1 Overall system architecture

The system architecture is not only suitable for large users' distributed deployment, acquisition and processing needs, but also can meet the needs of single user centralized deployment. When the user only needs a subsystem, the interface file can be configured to meet different user needs.

### 2.2 Web information collection subsystem

Web collector usually starts from the URL called seed and extends to other required pages on the web through protocol. Therefore, more than 30% of the URL collection efficiency is not suitable for

the topic collection in the URL collection system. At the same time, how to obtain effective web page information and filter the noise of advertisement and navigation bar will directly affect the performance of subsequent intelligent processing.

### **2.3 Information intelligent processing subsystem**

After a certain download task is completed, a message is sent to activate the processing subsystem. The system will automatically process the downloaded content, including automatic classification, automatic summary and metadata analysis, such as creating text title, keyword, analyzing author, etc.

### **2.4 Information release subsystem**

In recent years, more and more attention has been paid to the form of information release. As an external information service platform, the main features of the subsystem are: multi perspective, multi-level release and collection of information, that is, cross display from multiple perspectives of source, original column and classification system, which can flexibly carry out the intersection and combination of information sets; personalized information release, After logging in, users can use the personalized information customization interface to select the information perspective they are interested in. After logging in again, the user will be pushed with completely personalized information content. The powerful information retrieval ability not only provides retrieval for independent fields, but also provides fast retrieval, advanced expression retrieval and full-text retrieval.

## **3. Analysis of university library information management process**

University library is an academic service institution which provides information services for information users inside and outside the university, and its core business is to build library collection resources and provide information services (including tangible and intangible information services). In this paper, the organization's process management will be analyzed from two aspects, namely, modern collection management (logistics management) and information flow management.

### **3.1 Modern collection management**

Compared with logistics management in enterprise management, collection management can be regarded as the "logistics" management of library. Compared with collection resources in the traditional sense, modern collection resources have great changes in structure, including non-digital collection and digital collection, and the corresponding collection process management is also being adjusted and changed. Modern collection management refers to a series of management processes, including collection, purchase, storage, processing, circulation, reading, transmission and communication of traditional non-digital collections as well as modern digital collections. The modern collection management process is analyzed in Figure 1.

Collection management aims to construct, develop and utilize collection information resources, and its information activities can be divided into two stages: one is the formation stage of information resources, namely, a series of activities such as collection, purchase, storage, processing and display of collection resources; the other is the development and utilization of information resources, including circulation, reading, transmission as well as exchange of collection resources, in order to realize the value of information resources and achieve the goal of information management.

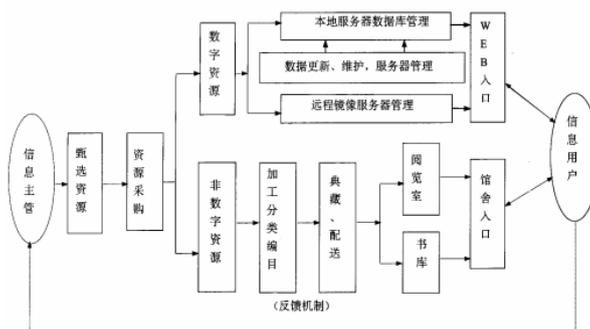


Fig.1 Collection management flow

### 3.2 Information flow management

The concept of library information flow is H1, which is generated by the application of computer and network technology in the library. It refers to the operation of intangible resources. All operations are generated by computers and achieved through the network, and manifested as information processing, storage, transmission, access, management, etc. Meanwhile, it has a close relationship with logistics, but the information flow has its unique rules. Some of the objects processed in the information flow are transformed from logistics, while most of them come from the information sources in databases and networks. With the improvement of automation and network technology, the proportion of online information resources will be larger and larger. Information flow management is divided into two levels, namely, resource information flow management as well as service information flow management. The information flow of resources refers to the information flow generated during the operation of library collection resources (including traditional collections and digital collections) conducted by internal and external staff. Specifically, in view of the library management information system, when the internal staff process resources, the internal resource information flow is produced. For external each user, if the information demand is different, the information access behavior will be different, and he (she) will leave user log information flow in the process of selecting, borrowing books, and retrieving digital resources. Service information flow management is a kind of information service management based on user information consultation, and it is the second level of information flow management based on resource information flow management as the first level. Information technology has brought vast information ocean, accompanied by information overload, that is, people's ability to create information has exceeded their ability to retrieve and process information. Information consulting service is the process of processing, sorting and combining various information resources in the collection based on users' requirements, and providing users with information solutions. Service information flow is the process management of processing, organizing and transmitting all kinds of information produced by information consulting service activities.

### 4. Overall scheme design and implementation

Based on the above analysis of organizational process management, the author believes that the construction of WEB information space of university libraries based on the concept of IA should include the following five modules: WEB site, public retrieval system (OPAC), WEB information management system, virtual consulting service system, as well as information feedback and evaluation mechanism, as shown in Figure 2. The technical realization of library collection logistics management depends on the public retrieval system, while the technical realization of information flow management relies on the WEB information management system, virtual consultation system, and information feedback and evaluation mechanism. The above-mentioned subsystem modules are integrated on the platform of WEB sites, in order to provide a unified WEB entrance for users.

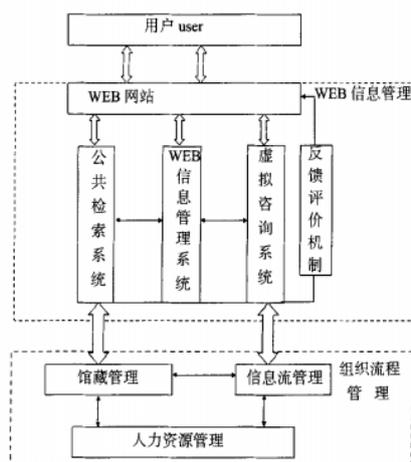


Fig.2 WEB information management system design

#### 4.1 WEB site design

The design of WEB site is divided into several stages, including goal programming, user analysis, modeling design, and prototype testing. First of all, the organizational goals of the WEB site should be determined, and a definite and clear position should be provided for the organization to build the WEB site. The purpose of university library WEB site is to reveal library collection and provide services, and provide an information service platform based on WEB for users to access and communicate information. Therefore, the design orientation of WEB site location must adhere to the “people-oriented” service concept, focus on users’ information trend and demand, and provide information function services, such as convenient and fast information query and information navigation. Secondly, the user groups served by WEB sites should be made clear, and their goals and requirements should be understood. Meanwhile, users’ WEB operation skills, basic configuration of computer, network bandwidth, WEB browser and other factors should be considered when designing WEB page. In the design modeling stage, the information content of WEB site should be started first, and then the WEB system and page frame structure should be designed, and finally, the interaction design of the WEB interface should be completed.

In fact, WEB interaction design is a process to realize the communication and communication between the organization’s WEB information and WEB users, rather than a technical implementation process of simple WEB page interaction, so the WEB interaction process is characterized by multi-orientation, multi-latitude as well as effectiveness. Based on the information demand characteristics of university library WEB site users, the author believes that, on the one hand, the organization of information content should be marked prominently and expressed orderly, in order to effectively convey as much information content as possible to users within a limited display pixel range. On the other hand, the design interface should be fresh and lively, and the page frame structure should be simple and smooth. At the same time, there should not be too many link levels, so as to avoid the use of unnecessary pictures and dynamic effects, and improve the WEB response speed and download time. Finally, the process of prototype testing is essential to maintain the vitality of the site. To meet the constantly changing information needs of users, a reasonable WEB site user evaluation and feedback mechanism must be established to constantly update and improve the WEB site.

#### 4.2 WEB information management system

The construction of WEB information management system is substantially to establish a classification database for all published information of the website and realize interactive dynamic operation based on the form of WEB. The website management can realize the dynamic addition, deletion and classification of information management, and when seen from the perspective of customer use, it can conveniently and fast inquire related topic information. According to the content published by WEB information

### 4.3 Virtual consulting system

Virtual consulting service is the development direction of digital library user service, and it is an intelligent service provided for virtual users under the support of virtual consultants [5]. The virtual consultation service mode based on WEB has diversified forms and different characteristics. According to the realization form of WEB page, consulting service can be divided into static consultation and dynamic consultation, and according to the mode of information transmission and receiving, it can be divided into active consultation and passive consultation. In this paper, ASP technology is used to design and implement WEB virtual consulting system. Figure 3 shows the data flow diagram of the system. As can be seen from the master-control interface, the sub-functions realized by the system include login, password modification, information consulting, information inquiry, consulting management, user management, user inquiry, and login again, each of which establishes its own sub-module, thus realizing the dynamic interactive platform of information user online consultation.

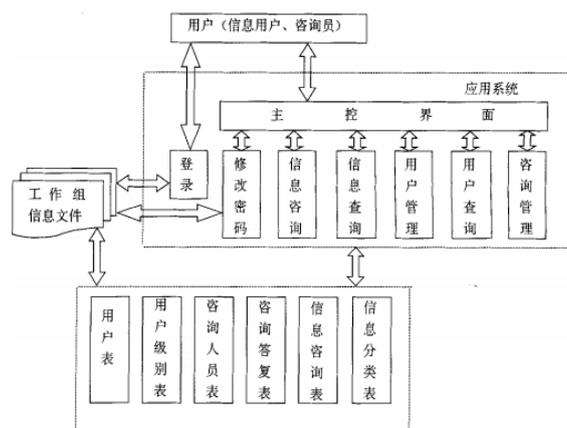


Fig.3 Virtual reference system data flow

### 5. Functional analysis of overall design

In order to improve the effectiveness of the download page, it is necessary to filter out some useless noise information. Traditionally, there are two solutions: one is to extract the page block with the largest entropy value by page learning; the other is to extract the interested content that needs to be accessed by defining the access path. Either scheme must have the ability to parse the HTML of the page, that is to construct appropriate data structure to parse HTML tags. The difficulty of parsing lies in the fact that the HTML of web pages is not standardized and strict. Even if there are errors or loose matching between tags, ie can also present perfectly. However, in order to extract information accurately, it is necessary to construct a strict access structure. Abstract is a process that extracts the important information to meet the needs of users by processing the content of the document, and then generates a more refined abstract than the original text after reorganization and modification. At present, there are three main types of automatic summarization technology: the method based on shallow analysis, the method based on entity analysis and the method based on discourse structure.

### 6. Conclusion

Information construction is a new concept of information management. The purpose of designing web information management system based on the concept of information construction is to build clear and usable web information space for users. It is not only an important web service platform provided by University Library for users, but also an important part of building a good information ecological environment. Therefore, it is of practical and practical significance to study and design the web information management system of university library Righteousness.

## References

- [1] Yin Cunyan, Dai Xinyu, Chen Jiajun. Automatic text summarization technology on Internet. Computer engineering. Vol.32.no.3.feb.2006
- [2] Jiang Yongchang. Scientific methodology of knowledge construction research based on knowledge transformation mechanism [J]. Information theory and practice, 2015, 38 (4): 11-17
- [3] Jiang Ning, Wang Weixia. Value creation of customer information experience and construction of information competitive advantage in supply chain under Web3.0 environment [J]. Information theory and practice, 2015, 38 (8): 79-84
- [4] Hu Haibo. Information aggregation service based on user interest in Web 3.0 environment [J]. Intelligence theory and practice, 2014, 37 (8): 117-121
- [5] Liu Qiang, Zeng Minzu. Information construction system and its influence on promoting the progress of information service industry [J]. Information theory and practice, 2003, 26 (1): 1.7