

The Construction of the Resources Recommendation Platform in University Libraries

Jingjing Liu^{1,*} and Jianxin Tian²

¹Library of Beijing Institute of Fashion Technology, 2 East Sakura Road, Chaoyang District, Beijing, China

²Beijing Institute of Fashion Technology 2 East Sakura Road, Chaoyang District, Beijing, China

*Corresponding author

Abstract—On the basis of a comprehensive analysis of the studies on the resources acquisition pattern in university libraries and Patron-Driven Acquisition at home and abroad, this paper focuses on the design and development of online recommendation system of university libraries resources, with the technology of the analysis of user requirement information, the duplicate checking of resources and auto-push, and the development of the mobile platform. The final decision on the acquisition and construction of library resources is given to the whole school so as to resolve the contradiction between the limited knowledge structure of acquisitions librarians and the huge amounts of literature content, which will also make the acquisition more specific and practical and reduce the blind and random selection of resources. As a result, the quality of the construction of literature resources and the utilization of literature will be improved.

Keywords—driven; resources recommendation platform; resources acquisition; readers' decision making

I. INTRODUCTION

Patron-Driven Acquisition (PDA), an emerging technology and research field of the rapid development of the library field in the network environment, has been one of hot research topics of many famous universities and research institutions at home and abroad in recent years[1]. It is an E-book business model raised by Net-library, a brand new book acquisition pattern with books ordered based on needs, different from the traditional one, and a new concept of the construction of library collections. This acquisition pattern plays an important role in resolving the common supply and demand contradiction of literature at present and improving the utilization of literature. Its core idea is to believe that readers have a greater say than acquisitions librarians or subject librarians. It is readers who decide on the types of E-books to be bought, which will make the most of purchasing funds. It is also the major pattern for many university libraries to buy books, especially E-books, so it is regarded as one of ten tendencies to have an impact on university libraries by ACRL in 2012[2].

At present, there are many foreign university libraries studying and practicing PDA; however, most domestic libraries adopt "reader recommendation" pattern, which is in accord with our country's present situation and a transition to complete acquisition based on readers' decision under present conditions. It is a more feasible pattern and approach satisfying the present features of manpower, funds, users, bibliography

and environment, which has good prospects for application and popularization. There are mainly two types of recommendation systems in university libraries: one is the automatic integrated management system including reader recommendation module. But it has been found that the recommendation modules are, to some extent, too simple to meet the need of practical work while being used. The other one is the professional software developed especially for reader recommendation, such as the Chinese book recommendation system developed by Wuhan University, which offers the service of receiving and processing the suggestions of Chinese book recommendation, Weblink developed by Taiwan Bookhouse Culture and Science Technology Corporation, PSOP developed by China National Publications Import and Export Corporation and the Chinese E-book Platform of Imagination Star developed by Rentian Book Corporation[3]. All of these softwares have the connecting and configuration problems with library portal websites and the back-end data, which need local secondary development and construction. So, it's indispensable to develop an online recommendation system and service operating mechanism appropriate for one's own library by collecting extensively the requirement information of university users' books and database and setting up requirement database accordingly, analyzing user behavior with intelligent models and being clearer about readers' requirement dynamic.

II. USER REQUIREMENT ANALYSIS

User requirement is the foundation and ultimate appeal of all the work. So, first of all, it's necessary to do the investigation and statistical analysis of the reader requirements, that is, that of our whole school. The user requirement analysis intends to be made in detail: user type (user's identity, major, reading habit, source, etc.), requirement resource type (paper resources, electronic resources, books, journals, newspapers, pictures, information, videos, etc.), access (receiving emails, SMS reminder, instant response from the platform, etc.)[4]. Meanwhile, the library requirements should also be studied in detail, such as user statistics, resources statistics, the assessment of librarians' workload, which will make sure that the design and development of the platform can meet various needs.

Besides, set the back-end and interface parameters on the basis of fully considering conciseness and practicability, the analysis of user and library requirements and the design of functional modules. It intends to include: recommendation

privilege, administrator privilege, ordinary user privilege, advanced user privilege, statistical distribution of the amount of money for recommendation, response time and fault tolerance.

III. THE DESIGN OF THE GENERAL FRAMEWORK

On the basis of previous investigation and analysis, set up an intelligent analysis model of user requirements, which intelligently analyzes user requirements with the technology of big data analysis. SQL SERVER and .NET are also used to develop the system platform, including the design of the front page, the demonstration of catalogue at all levels, and the design and development of functional modules, such as the verification module of back-end readers, the recommendation module of subscription catalogue, auto-recommendation module, the automatic order processing module and the management module of reader comments. Finally, the library website will be embedded in it.

The library resources online recommendation system intends to apply the design of B/S three-tier architecture, whose main part contains three layers: user interface, business logic and data, including the development of environment platform, development tools, database management system, statistical analysis tools and version control tools. The focus of the study and development is on the interface between online recommendation system and library book retrieval system, the analysis and mining technology of the resources of reader recommendation, and the recommendation of electronic resources, especially database resources.

It mainly includes the construction and development of subscription catalogue database at the data layer, personal information management system, site retrieval technology, recommendation management system and system management technology, which are at the business logic layer, and the study on the implement method of WeChat library and online recommendation platform.

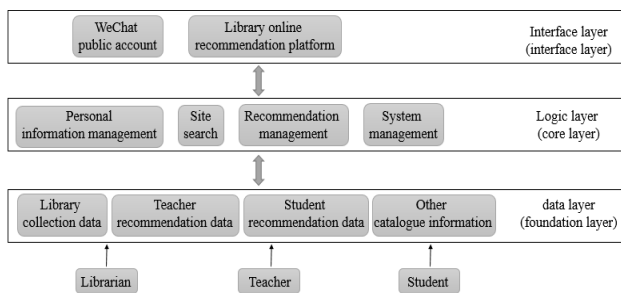


FIGURE I. OVERALL BLOCK DIAGRAM OF RESEARCH SCHEME

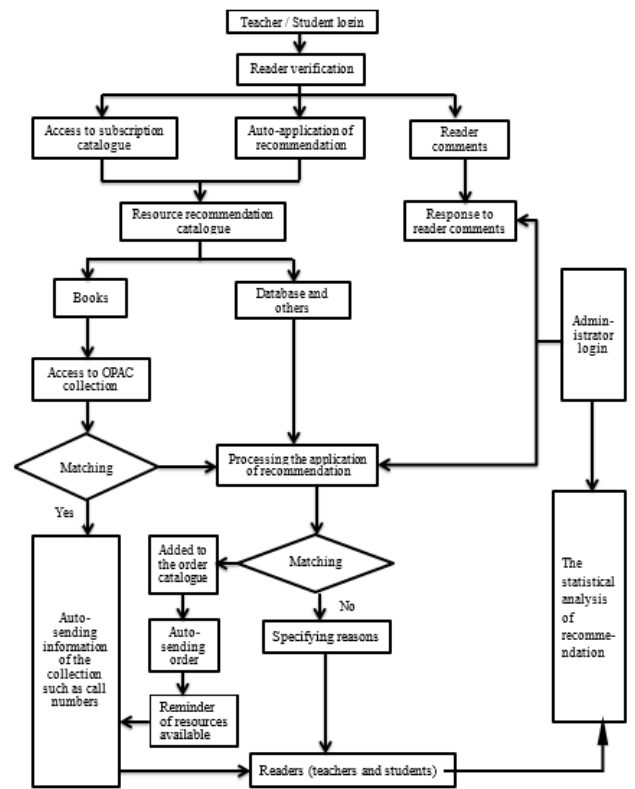


FIGURE II. THE DIAGRAM OF RESEARCH METHODS AND TECHNOLOGY ROADMAP

IV. THE DEVELOPMENT OF SUBSCRIPTION CATALOGUE DATABASE

The subscription catalogue database, based on SQL Server, will be developed according to the data from publishers, suppliers of books and database and Douban. One important factor to construct the subscription catalogue database is to set up the classification system. The classification systems are different in various industries, but their forms are basically the same by applying multi-tier classification, with each one consisting of name and code. The subscription catalogue database includes several functions as follows:

- to support the management and construction of multi-tier classification
- to support the addition of the classification names and codes at all levels
- to support the import and export of the classification system
- TO SUPPORT THE INDEXING OF MULTI-CATEGORY.

V. THE STUDY ON THE DUPLICATE CHECKING OF RECOMMENDATION RESOURCES AND OPAC COLLECTION RESOURCES AND AUTO-PUSH

As for the library resources recommendation platform, the important and difficult part lies in the development of the duplicate checking of collection resources and the auto-push for different users. According to the content of user recommendation, the duplicate checking can be divided into

three categories: simple duplicate checking, the duplicate checking based on the content and the full-text duplicate checking. The simple duplicate checking applies the database search, checking the information of book titles and ISBN, etc. by matching the data fuzzily. The full-text duplicate checking uses the full-text retrieval on the basis of present word-separation technology.

As for the duplicate checking based on the content, users upload the books' covers and inserts which they are interested in and, according to the image content, do the duplicate checking with the content-based image retrieval technology. This project intends to apply the algorithm of image feature retrieval based on Surf. The extraction and detection of the natural feature of the image has always been the hot research topic in computer graphics. There are many kinds of algorithm about the extraction of natural features in recent years, such as Sift(Scale Invariant Feature Transform), Surf(Speeded-up Robust Feature), Fast(Features from Accelerated Segment Test), Brift(Binary Robust Independent Elementary Features), etc[5]. Sift and Surf are the feature detection algorithm of scale space, with the scale not changed. Surf, raised by Bay et al, is an improved Sift. It is mainly divided into three parts:

a. feature point extraction. The Hessian of the feature points from which to choose and other points will be calculated based on the integral image by replacing the second order Gaussian filter with the box filter. The highest one is the feature point.

b. feature point description. Haar will be calculated around the feature point, so do the four sums, which constitute the feature description.

c. matching with the vector described by the feature point. The most distinguishing feature is the application of Harr and the integral image, the replacement of the second order Gaussian filter with the box filter, the acceleration of convolution with the integral image, which will reduce the complexity of calculation time and increase the calculation speed. The flow chart of Surf is as follows in figure 3:

Surf is also improved according the features of book information and mobile terminal, which will make it more suitable for the feature of the limitation of mobile device resources, and more efficient with the premise of ensuring correctness. The following three aspects are included:

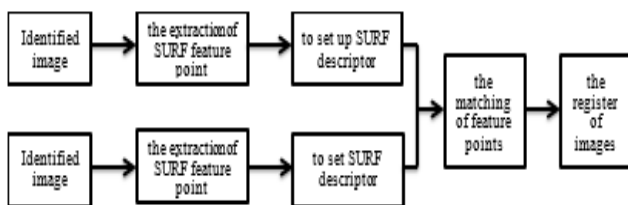


FIGURE III. LOW CHART OF SURF ALGORITHM

A. The Dynamic Building of Image Pyramid

In order to speed up the detection of feature points, the number of building image pyramid layers in Surf is related to the length and width of the image, which realizes the building of pyramid layers in a dynamic and self-adaptive way, and

improves response time as well. The minimum of the length and width of the image will be used to get the number of pyramid layers, as in formula (1):

$$O = \frac{\ln S}{C} - a \tag{1}$$

In formula (1), S is the minimum of the length and width of the image; C is the constant; is the threshold value, which ensures the scope of the number of pyramid layers. C, in this project, is 1.5 and is 2.

B. Feature Points Filtering

There are many feature points in Surf, unevenly distributed, which will lead to the local optimum. During the distance constraint, the most optimal feature point might not be in the most optimal set, which will fall in the local optimum. So, with Surf improved, the feature points will be filtered first. The Harr corresponding value within a certain radius of feature points will be calculated. The feature points whose corresponding value is less than some threshold value will be filtered. Some identical ones will also be taken out with niche from genetic algorithm.

The formulas to calculate the distance ()between two feature points are (2) and (3):

$$\|x_i - x_j\| = \sqrt{\sum_{k=1}^s (x_{i_k} - x_{j_k})^2} \quad \begin{matrix} i = 1, 2, \dots, M + N - 1 \\ j = i + 1, \dots, M + N \end{matrix} \tag{2}$$

$$\bar{D}_{ij} = \frac{\|x_i - x_j\|}{S} \tag{3}$$

If $0 < \bar{D}_{ij} < \varepsilon$ (ε indicates the scope to decide whether they are identical, whose specific value will be set according to the situation), the two ones are identical. The Harr corresponding values of two feature points are to be compared, the smaller one of which will be put in a penalty function in order to decrease the feature corresponding value[7]. The penalty function in this project is:

$$f \min(x_i, x_j) = Penalty$$

C. Improving Fast Matching

In order to speed up the feature matching, improved BBF is applied, which will reduce the amount of calculation. While the Hessian determinant of a matrix is being calculated, the Hessian trace of matrix t is also calculated. If t's value is more than 0, the attribute of the descriptor is "plus"; if it is less than 0, the attribute of the descriptor is "minus". The feature points of each image will be divided into two groups according to the value of the descriptor and the data of the same type will be

compared. If the attributes of the two descriptors are not the same, the feature points are not matched and won't be compared any more.

As for the mal-matching problem among the matched feature points, the wrong ones will be taken out with RANSAC. Then the first n matched featuring points with the highest similarity will be selected according to the program. The precision of matching will be achieved by assigning n.

Push and pull are two ways to get information at present[8]. Each of them has its own advantages and disadvantages. This project applies "Intelligent Information push-pull (IIPP)" instead to push recommendation resources to the users flexibly. IIPP, applying some methods, such as artificial intelligence, machine learning, knowledge discovery and reasoning, etc., improves the "information source's level of speculation about users' interest and achieves the auto and personalized push[6]. Based on WeChat open port and dynamic adaptation principle, the perception engine is constructed at the client due to the resource limitation of mobile terminals, perceiving the context. When there is message to be pushed, the message will be sent to strategy generation engine and the strategy will be generated, which will be sent to behavior engine and push will be completed.

As for some repository recommendation, artificial retrieval and response will be applied according to the present situation. This project intends to introduce the form processing flow in ERP and process the database resources and push information artificially, which will digitize and normalize the recommendation system flow for future extension of the automatic flow.

VI. THE STUDY ON USERS' BEHAVIOR AND ANALYTICAL AND MINING TECHNOLOGY OF RECOMMENDATION RESOURCES

The users' behavior and recommendation data will be analyzed and mined based on the construction of information database of users' requirements. Data mining refers to the mining and extraction of unknown and valuable knowledge from the huge amounts of data. The data analysis structure applied in this paper is as follows:



FIGURE IV. DATA ANALYSIS STRUCTURE

VII. CONCLUSION

The major study of this project is Patron-Driven construction of library online recommendation system and WeChat recommendation system, that is, with the concept of PDA, the initiative of library resource acquisition and resources construction is turned over to users. The internet-based reader recommendation system and WeChat-based instant recommendation system are developed, which build a professional platform on which readers take participate in the construction of library resources. On this platform, readers can

set up their own accounts, manage personal information, access to library's resources information, resources dynamic and other catalogue information of all kinds and then find the resources needed. They can also make recommendations themselves and express their own needs and wishes. It is not only a platform for bibliography information, but also a visual acquisition environment which is interactive, open and possible to communicate and interact. The problems that readers can't get the needed resources and the acquisition in libraries is aimless and at random have been solved. The utilization of library has been improved. The maximum of buying resources with limited fund has been achieved. The project is innovative and practical.

ACKNOWLEDGMENT

In this paper, the fund project of Beijing Municipal Education Commission (KYJH02160201/002/022) "based on the research of university library resources recommended user driven purchase platform" and the Beijing University Library Research Fund Project (BGT2016026), the key project of Beijing Institute of Fashion Technology (ZDJG-1611) is one of the research results.

REFERENCES

- [1] Zhang Dan, Xiao Ping. Considerations on the controversial issues in the implementation of reader driven procurement (PDA). *Library magazine*, 2012 (7): 24-27.
- [2] NISO. NISO publishes updated recommended practice on SERU: A shared electronic resource understanding [EB/OL]. [2013-03-25]. https://www.niso.org/news/pr/view?item_key=922f576ca5099b17ad8438a7bf8e0858f36430f3.
- [3] ACRL. 2012 top ten trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education [J]. *College & Research Libraries News*, 2012, 73(6):311-320.
- [4] Wang Xiaoping J of university library. Enlightenment reader recommendation _ problems and Countermeasures Caused by foreign PDA mode. *The library*, 2015 (1): 76-78.
- [5] Chen Daqing. Review of the development of e-book and future J. *Library Journal*, 2014 (5): 84-91.
- [6] Zhou Cong. University Library recommendation system research and design of J. *Journal of library*, 2014 (12):51-54.
- [7] Liu Jun, Jin Shuna. Research on reader decision making based on knowledge discovery [J]. *Library science research*, 2014 (1): 42-45.
- [8] Lai Group. An empirical study of reader decision making and purchasing (PDA) model in book acquisition of higher vocational library [J]. *Library sector*, 2013 (3): 74-77.