Research on Constructing User Profiles for University Libraries Based on Large Language Models[*]

Wang Yibo¹ Zhang Pengyi²

2026-07-24

ChinaRxiv

Abstract

The construction of user profiles is the foundation for analyzing user characteristics and understanding user needs. This study proposes a method for constructing and updating user profiles in university libraries based on large language models and data mining technologies. First, a tag system for university library user profiles is constructed, including 5 first-level indicators such as user attributes and user behavior, 11 second-level indicators such as personal information and academic background, and 49 third-level indicators such as name, department, time spent in the library, and borrowing frequency. Second, methods for user profile construction, updating, and profile generation based on large language models are proposed. Finally, taking readers of Peking University Library as an example, an empirical study verifies the effectiveness of the proposed method, designs dynamic individual and group user profiles generated based on large language models, and realizes the visual display and dynamic updating of user profiles.

Full Text

【Technology ·Platform】

Research on Constructing User Profiles for University Libraries Based on Large Language Models*

—Taking Peking University Library as an Example

Wang Yibo1 Zhang Pengyi2

1. Peking University Library, Beijing, 100871
2. Department of Information Management, Peking University, Beijing, 100871

Abstract The construction of user profiles is the basis for analyzing user characteristics and understanding user needs. This study proposes a method for constructing and updating user profiles in university libraries based on large language models and data mining technologies. First, a tag system for user profiles in university libraries is constructed, including five first-level indicators such as user attributes and user behaviors, 11 second-level indicators such as personal information and university background, and 49 third-level indicators such as name, department, length of stay in the library, and borrowing frequency. Second, methods for user-profile construction, updating, and profile generation based on large language models are proposed. Finally, taking readers of Peking University Library as an example, an empirical study verifies the effectiveness of the proposed method. Dynamic individual and group user profiles generated based on large language models are designed, enabling visual presentation and dynamic updating of user profiles.

[**Keywords**] large language model; artificial intelligence; university library; user profile; data mining

[**CLC Classification Number**] G252 [**Document Identification Code**] A [**Article Number**] 1003-7845(2026)02-0057-08

[**Citation Format**] Wang Yibo, Zhang Pengyi. Research on Constructing User Profiles for University Libraries Based on Large Language Models—Taking Peking University Library as an Example[J]. *University Library Work*, 2026(2): 57–64.

Introduction

With the continuous development of the era of big data and artificial intelligence (AI), the library field has gained unprecedented development opportunities. Against this background, AI-empowered university libraries should not remain limited to traditional basic functions such as book lending and document management, but should gradually transform toward user-oriented intelligent knowledge services. This transformation requires libraries not only to uphold a service philosophy centered on user needs, but also to identify users' personalized needs accurately. University libraries possess abundant collection-resource data and user data. As a technical method for analyzing user characteristics and behavioral patterns, user profiling can collect and analyze data such as users' demographic information, behavioral characteristics, and preferences from multiple dimensions, thereby constructing concrete and vivid user profiles and comprehensively presenting the behavioral characteristics of different groups.

However, existing studies mainly focus on the theoretical construction of user-profile tag systems, methodological discussions, and service models, and they generally rely on expert judgment and clustering methods in traditional machine learning[1]. On the one hand, because users' interests and needs have characteristics of dynamic evolution, traditional static user profiles are unable to reflect changes in real time. On the other hand, existing research has insufficiently explored the practical application of frontier technologies such as knowledge graphs and large language models, and in particular lacks empirical research integrating multi-source data. Based on this, by collecting and analyzing real data such as users' basic information, borrowing behavior, and library entry-exit records, this study conducts modeling analysis of user data, uses technologies such as large language models and data mining to realize the automatic generation of user profiles, and proposes a dynamic updating method for user profiles.

1 Current Status of Related Research

A user profile is a virtual representation of a real user, used to describe the typical characteristics of a target user; it is a concrete form of presentation of a user model. This concept was first proposed by Cooper A, an expert in the field of interaction design[2]. Baxter K et al.[3] further defined a user profile as a detailed description of user attributes, including job title, experience, educational level, key tasks, age range, and other content; it is a multidimensional and detailed set of features. In essence, a user profile is a structured user model formed on the basis of user data, describing users' basic attributes, interests, preferences, and other characteristics through tagging[4]. Traditional methods for constructing user profiles mostly adopt expert argumentation: experts predesign a user classification system and rules, obtain user information through questionnaires, interviews, and other methods, and then form user profiles through manual analysis and induction[5]. In recent years, in addition to e-commerce, libraries, and other

---

* This article is one of the research outcomes of the National Key R&D Program project "Innovation in Smart Library Theory and Methods in the Era of Artificial Intelligence" (Project No.: 2023YFF0906801).

Author profiles: Wang Yibo, PhD, librarian; research interests: data mining and analysis, data management and services. Zhang Pengyi, PhD, tenured associate professor; research interests: user behavior analysis, information and knowledge organization.

Received: 2025-12-10 **Responsible editor:** Yang Qianzi

In addition to being applied in depth in these fields, user profiles have gradually expanded into biology [6], architectural science [7], and other domains; different disciplines place different emphases on the understanding of user profiles. In the areas of recommendation algorithms and recommender systems, a user profile is a data-driven visual representation of user behavior and preferences. For example, Godoy D et al. [8] describe a user profile as a record of user behaviors, interests, and characteristics, which can be obtained through questionnaires and interviews or generated with the aid of machine learning and data mining. Such user models form the basis of personalized information retrieval and recommender systems; systems can provide content that meets user needs on the basis of user profiles, thereby improving user satisfaction.

At present, the practical application of library user profiles is in an active exploratory stage. They have been incorporated into diverse service scenarios such as electronic course reserves, manuscript recommendation, and subject consultation. While optimizing readers' actual experience, they further enhance the library' s value dissemination capacity and social influence. Relevant studies include designing context-configuration schemes for different library application scenarios, reconstructing multidimensional user profiles, developing user-profile visualization platforms, improving the effectiveness of personalized retrieval in digital libraries, optimizing system functions and interface design, and providing scenario-based recommendations for different users. In terms of overseas practice, university libraries such as Stanford University, Liberty University, and North Carolina State University have already carried out user-profile-related practices [9]. In domestic practice, the National Library of China, Hubei Provincial Library, Guangxi Zhuang Autonomous Region Library, Shenzhen Library, Jiaxing Library, and others are also actively advancing intelligent platform construction and promoting the implementation of user profiles in library service scenarios. However, at the present stage, the practical application of library user profiles is still constrained by the level of platform and service intelligence, and both the breadth and depth of application need to be improved. The main limitations are as follows: the profiles generated by traditional methods are mostly discrete keywords or numerical tags and lack natural-language descriptions; profiles have not yet achieved multimodal presentation forms, such as AI virtual portraits capable of reflecting the characteristics of individual or group users; after static profiles are constructed, they are not updated in a timely manner and therefore cannot promptly reflect changes in users' behavioral characteristics, interests, and preferences.

By contrast, AI-enabled user-profile construction methods introduce technologies such as machine learning and large language models, which can help automate and intelligentize profile construction. Scholars have begun to use data-mining algorithms to extract and mine user characteristics from logs—for example, dividing users with similar behaviors into groups through cluster analysis and mining association items of user preferences through association rules. User behavior data are converted into feature vectors, and algorithms such as clustering, regression, and classification are combined to automatically complete user-profile construction and feature-tag extraction. Compared with manual construction methods, machine-learning methods can discover and summarize user characteristics from implicit data and, relying on continuously added data, realize dynamic updating of profiles. In recent years, deep-learning technologies have also been gradually applied to user-profile construction. Models based on deep neural networks can automatically learn low-dimensional vector representations of users and resources, thereby generating more accurate user profiles.

Against this background, this study proposes a method for constructing and updating university library user profiles based on large language models. Its main innovations include: first, constructing a profile tag system that incorporates generative indicators, and using large language models to transform structured data into natural-language descriptions and personalized virtual avatars, thereby realizing multimodal presentation of profiles; second, proposing a tag-extraction and dynamic-updating method based on the RFCI model and text mining, providing a reference for the intelligentization and personalization of knowledge services in university libraries.

2 Construction of University Library User Profiles Based on Large Language Models

2.1 Design of the User-Profile Tag System

This study intends to construct a user-profile tag system that integrates the generative capabilities of large language models. Two key factors need to be considered: first, satisfying the need for automatic updating of user profiles and promptly identifying dynamic changes in user interests and behaviors; second, adapting to the characteristics of large language models so as to realize the automatic generation of some tags for individual and group profiles. On the basis of reviewing relevant domestic and international studies [10-16], and comprehensively considering various types of user-profile tags, this study determines a basic profile framework comprising "static attributes" and "dynamic behaviors," and constructs an initial multidimensional user-profile tag system that includes four first-level indicators: user basic information, user usage behavior, user preference characteristics, and user group comparison.

To further improve the indicator system, six experts and scholars in the discipline of information resource management with rich experience in the field of user profiles were invited—one full senior professional, two associate senior professionals, one intermediate professional, and two doctoral students—to evaluate the structural completeness of the indicator system and the scientific validity of the indicators. The indicator system was improved according to expert opinions: based on the generative capability of large language models in the AI era, the first-level indicator "user description" was added; tag names were standardized, for example, the original "user basic information" was renamed "user attributes," and "user usage behavior" was adjusted to "user behavior." Ultimately, a university library user-profile tag system was constructed, including five first-level indicators: user attributes, user behavior, user preferences, group ranking, and user description. Under these are 11 second-level indicators, such as personal information, university background, library-entry behavior, and use of print resources, which are further refined into 49 third-level indicators, such as name, department, library-entry frequency, and borrowing frequency, as shown in Table 1.

2.2 Methods for Constructing User Profiles

User attributes are the basic data for constructing user profiles. However, because student/staff numbers, names, and similar information involve personal privacy, this study applies corresponding processing through a general coding-rule replacement method, symbolic transformation, and desensitization of the name field.

In constructing user-behavior tags, this study refers to the logic of the RFM model and constructs an RFCI model. Here, R is defined as the time-interval indicator for the user's most recent use of a service, F is the service-use frequency indicator, M is replaced by the user cycle C, and the indicator I is added to represent the degree of dispersion or concentration in service use; thus, the RFCI model is constructed. This model is applicable to data analysis of users' library-entry behavior ($B1$), use of print resources ($B2$), library-entry performance ($D1$), and performance in the use of print resources ($D2$), with the aim of

The main user behaviors are digitized and presented in the form of labels. The specific definitions of user behavior-profile labels are shown in Table 2.

Table 1 Library User Profile Label System

First-level indicator	Second-level indicator	Third-level indicator
User attributes (A)	Personal information (A1)	Name (A11), gender (A12), age (A13), email address (A14)
User attributes (A)	Campus background (A2)	Student/employee ID (A21), identity (A22), school/department (A23), major (A24), year of enrollment/employment (A25), grade (A26, for students), graduation year (A27, for students), educational background (A28, for students), research field (A29, for teachers), post/professional title (A210, for teachers)
User behavior (B)	Library-entry behavior (B1)	Library-entry date (B11), library-entry frequency (B12); days in the library (B13), duration in the library (B14)
User behavior (B)	Use of print resources (B2)	Borrowing date (B21), borrowing frequency (B22), borrowing time period (B23), borrowing duration (B24); return date (B25), return frequency (B26), return time period (B27), return duration (B28); reservation date (B29), reservation frequency (B210), reservation time period (B211), reservation duration (B212)

First-level indicator	Second-level indicator	Third-level indicator
User behavior (B)	Use of electronic resources (B3)	Search volume in journal databases (B31), download volume from journal databases (B32); search volume in book databases (B33), download volume from book databases (B34)
User preferences (C)	Library-entry preferences (C1)	Preferred month for library entry (C11), preferred time period for library entry (C12)
User preferences (C)	Preferences in the use of print resources (C2)	Preferred borrowing month (C21), preferred borrowing time period (C22), preferred borrowing keywords (C23), preferred borrowing topics (C24), breadth of borrowed disciplines (C25), relevance of borrowing behavior (C26)
Group ranking (D)	Library-entry performance (D1)	Percentile ranking of number of library entries (D11), percentile ranking of days in the library (D12), percentile ranking of duration in the library (D13)
Group ranking (D)	Performance in the use of print resources (D2)	Percentile ranking of borrowed volumes (D21), percentile ranking of reserved volumes (D22)
User description (E)	User feature graph (E1)	A graph with feature labels for an individual/group user (E11)

User description (E)	Textual description of user features (E2)	A general textual description with feature labels for an individual/group user (E21)

Table 2 Definitions of Indicators in the RFCI Model for University Library User Behavior Profiles

Indicator code	Indicator meaning	Unit of measurement
JG_R	Within the statistical period, the number of days between the user' s most recent library-entry time and the observation endpoint	Days
JG_F	Within the statistical period, the number of times the user entered the library	Times
JG_C	Within the statistical period, the number of days between the user' s first library entry and last library entry	Days
JG_I	Within the statistical period, the concentration or dispersion of the user' s library-entry behavior; the ratio of the total number of library entries to the total number of days	Times/day
JY_R	Within the statistical period, the number of days between the user' s most recent borrowing or renewal time and the observation endpoint	Days

Indicator code	Indicator meaning	Unit of measurement
JY_F	Within the statistical period, the sum of the user’ s borrowing and renewal times	Times
JY_C	Within the statistical period, the number of days between the user’ s first borrowing or renewal and last borrowing or renewal	Days
JY_I	Within the statistical period, the concentration or dispersion of the user’ s borrowing behavior; the ratio of the total number of borrowing and renewal transactions to the total number of days	Times/day

Note: JG denotes library entry; JY denotes borrowing.

In constructing user-preference labels, this study adopts TF-IDF as the keyword extraction method; selects BERTopic as the topic-modeling method; uses the information-entropy method to analyze the breadth of disciplines in users' borrowing; applies association rules to identify associations in book borrowing

16

; and uses correlation analysis to explore the degree of association and behavioral similarity in borrowing preferences among user groups of different identities, genders, and educational backgrounds, thereby mining behavioral association features among user groups.

This study also comprehensively employs large language models to automatically generate user avatars and personal profiles. It selects the GLM-4.5 model released by Zhipu; this model has strong semantic understanding capabilities and is suitable for academic text generation. The user-feature portrait is implemented through GLM-4.5’s CogView-4 image-generation model. To standardize prompt design during the generation process, this study adopts the CO-STAR framework to construct prompts

17

.

2.3 Method for Updating User Profiles

In the process of constructing user profiles for university libraries, updating user-profile data is an extremely important component. User profiles must remain accurate, timely, and dynamic so that library services can meet users' constantly changing needs and interests. The main purposes of updating user-profile data include: reflecting dynamic changes in user behavior and preferences, so as to ensure that the current user profile can embody users' latest needs and

interest; the updated user-profile data can be used for model training and optimization; and continuously updating user-profile data helps improve user satisfaction and user stickiness. User attribute data are usually updated once per semester. The latest data are periodically obtained from the school management system, and users' background information at school is uniformly updated at fixed time points, such as the beginning or end of a semester. User behavior data are generally updated on a weekly or monthly cycle. Following the RFCI model method proposed above, users' library-entry behavior is recalculated, assigned values, and updated. User preference data are also updated synchronously on a weekly or monthly basis. For example, after the access-control system records users' time periods of entering the library, the proposed RFCI model method is used to regularly calculate in batches changes in user preferences in terms of length of stay in the library, frequency of library entry, and so forth. Rankings of users' number of library visits and length of stay among users in the same department and the same grade are updated regularly at a weekly or monthly frequency. If certain key user attributes—such as department, major, grade, or research direction—or borrowing preferences change substantially, a large language model can be invoked to regenerate the user feature graph; if user behavior and preferences change substantially, a large language model can be invoked to improve the descriptive content.

3 Empirical Study Based on User Data from Peking University Library

3.1 Data Acquisition and Preprocessing

To verify the effectiveness of the user-profile construction and updating methods described above, this paper conducts an empirical study using user data from Peking University Library as the sample. Stratified sampling was adopted for data collection, selecting students and faculty/staff of the Department of Information Management at Peking University who were enrolled or employed during the five-year time window from 2020 to 2024, so as to ensure the timeliness and disciplinary representativeness of the data. There are many possible combinations and classification criteria for constructing group user profiles, such as department, grade, major, and research direction. In this study, the combination of "user gender + user educational level + user identity" was selected to form group user profiles. The specific data include: user attribute data (415 records), including basic information such as year of enrollment and identity category. By identity category, there were 343 students in total (83%) and 72 faculty/staff members in total (17%). By gender, there were 207 males (50%), including 41 faculty/staff members and 166 students; and 208 females (50%), including 31 faculty/staff members and 177 students. Among the student group by educational level, there were 169 undergraduates (49%), 109 master’ s students (32%), and 65 doctoral students (19%). The user scale satisfies the minimum sample-size requirement for Spearman correlation analysis. Borrowing logs (13,661 records) and renewal logs (4,412 records), reflecting users’ resource-utilization behavior and preferences; library entry/exit logs (70,081 records), reflecting users’ spatial-use behavior and preferences.

3.2 Labeling User-Profile Data

3.2.1 User Borrowing and Library-Entry Behavior Labels For each user, the total number of library entries and the corresponding rank, as well as the total number of borrowings and reservations and the corresponding rank, were calculated; the percentile rank within the department was also computed. The selected feature data were standardized, the optimal number of clusters was determined using the silhouette coefficient method, and finally the clustering results were analyzed to divide all users into several user-profile types, as shown in Table 3.

Table 3 Classification of User Library-Entry and Borrowing Label Types

User-profile label	Features	Description	Number of samples/items	Percentage/%
Comprehensive demand type	High library entry, high borrowing	Users frequently visit the library and borrow a relatively large number of books, indicating strong overall demand for the library	57	13.73

User-profile label	Features	Description	Number of samples/items	Percentage/%
Space demand type	High library entry, low borrowing	Users mainly use the library' s space and facilities, while borrowing books relatively infrequently	9	2.17
Resource demand type	Low library entry, high borrowing	Users do not often visit the library in person, but borrow many books, indicating high demand for library resources	11	2.65
Potential development type	Low library entry, low borrowing	Users do not often use library services and are low-demand users	338	81.45

3.2.2 User Borrowing Keywords and Topic Labels The keywords of the books borrowed by each user and their TF-IDF values were calculated and statistically analyzed, generating keyword labels for each user to describe reading preferences. Table 4 shows the statistical results for some users' borrowing keyword labels.

Table 4 User Borrowing Keyword Labels and TF-IDF Values (Example)

Student/employee ID	Name	Keyword labels and TF-IDF values
19A10324	Zhang Liangliang	*A Song of Fire* (0.59), *Monogatari* (0.28), *Cheetah* (0.22), *Past and Present* (0.22), King Arthur (0.19), deep learning (0.17)
21B20114	Li Xiaoming	digital publishing (0.43), editing (0.26), library tower (0.26), China (0.21), publishing (0.17)

The titles of the books borrowed by each user were calculated and counted separately; all titles were aggregated, and after deduplication 7,689 distinct books were obtained. The BERTopic topic model was used to analyze and mine these 7,689 books, yielding 10 main domain topics in users' borrowed books, as shown in Figure 1.

Figure 1. Topic characteristics of users'borrowed books based on the BERTopic model

Based on the distributional characteristics of borrowing frequency by topic, the borrowing frequencies of each user for all topics were sorted in descending order. A threshold cutoff point x was then dynamically determined so that the cumulative frequency proportion of the top x high-frequency topics was greater than or equal to 60%. The set of high-frequency topics satisfying this condition was ultimately defined as the user' s borrowing-topic label based on the BERTopic model. Sample data are shown in Table 5.

Table 5. User high-frequency borrowing-topic labels based on the BERTopic model (example)

Student/Employee ID	Name	High-frequency borrowing-topic label
19A10324	Zhang Liangliang	Social Sciences and Communication Studies

Student/Employee ID	Name	High-frequency borrowing-topic label
21B20114	Li Xiaoming	Information Management and Network Technology Applications; Program Design, Programming Languages, and Data Analysis; Librarianship and Theoretical Methods

This paper represents the borrowing topics of user groups by adopting an aggregate calculation method: individual users are divided into different groups according to user attributes; the borrowing topics of all individuals within a group are aggregated; and the core topics of that group are extracted. The borrowing-topic labels of different user groups were obtained; sample data are shown in Table 6.

Table 6. Borrowing-topic labels of user groups (example)

Status	Gender	High-frequency borrowing-topic labels
Undergraduate	Male	Program Design, Programming Languages, and Data Analysis; Information Management and Network Technology Applications; Physics and Biological Research
Master’ s student	Female	Social Sciences and Communication Studies; Chinese Culture and Thought; Interdisciplinary Collected Works and Studies of Thought

3.2.3 Discipline-breadth labels for users’ borrowed books

Through information-entropy analysis, it was found that male undergraduates and male master’ s students have relatively high entropy values: the *Chinese Library Classification* subject categories of the books they borrow are diverse, and the range of subjects they cover is broad. By contrast, male doctoral students and female faculty/staff have relatively low information entropy: their borrowed books are concentrated in a small number of CLC classification numbers, and their subject distribution is relatively concentrated. In label classification, using a normal distribution based on the mean and standard deviation may lead to inaccurate results. Therefore, users with information entropy of 0—that is, whose borrowed books belong to only one first-level subject category in the *Chinese Library Classification*—were defined as the discipline-concentrated borrowing group. For the remaining users, a percentile-based method was used for classification: users in the first 60% percentile were set as the discipline-balanced borrowing group, and users in the latter 40% percentile were set as the discipline-broad borrowing group. According to the entropy-based discipline-breadth labels for the books borrowed by different user groups, the groups can be divided into three categories by ranking, as shown in Table 7.

Table 7. Discipline-breadth labels for books borrowed by user groups

Discipline-breadth label for borrowed books	User group
Discipline-broad borrowing type	Male undergraduates; male master' s students
Discipline-balanced borrowing type	Female undergraduates; female doctoral students; female master' s students; male faculty/staff
Discipline-concentrated borrowing type	Male doctoral students; female faculty/staff

In association-rule mining, the borrowing records were further processed by splitting each book into an individual item. After multiple experimental tests, 139 frequent itemsets and 90 association rules were obtained. Example results are shown in Table 8.

Table 8. Association rules for book borrowing (example)

Antecedent	Consequent	Support	Confidence
Foundations of Library and Information Science	*Information Resource Construction*	0.02	0.75
Elegant Conversations in the World of Books	*Chinese Classical Philology*	0.01	0.60

Antecedent	Consequent	Support	Confidence
Information Resource Construction	*Introduction to Library Science*	0.01	0.75

3.2.4 Correlation-based group labels for users' borrowing behavior

Spearman coefficient correlation analysis was used to analyze differences in identity and gender.

the similarities and differences in borrowing behavior among user groups. To eliminate interference from differences in the time students and faculty/staff spend on campus, the data window was limited to September 1, 2020, to July 1, 2024. A total of 10,592 borrowing and renewal records for users during this period were selected. Based on the 22 major categories of the Chinese Library Classification, statistics were compiled and box plots were drawn; the results are shown in Figure 2.

Note: The numbers in parentheses on the vertical axis represent the total sample size of books under that disciplinary category.

Figure 2. Box plot of the distribution of borrowing frequency for individual books in different disciplinary categories

As can be seen from Figure 2, the three book categories with the highest user borrowing volume are Culture, Science, Education, and Sports (562 books), Literature (183 books), and General Social Sciences (126 books). The distribution of book borrowing frequency differs significantly across disciplines. Books in History and Geography are very popular, but their overall borrowing frequency is relatively dispersed. For books in Philosophy and Religion, and in Politics and Law, both the median and the lower quartile of borrowing frequency are relatively low; however, these disciplines also contain a certain number of high-borrowing outliers. The borrowing-frequency distributions of books in Medicine and Health, as well as in Astronomy and Earth Sciences, are highly concentrated, with no obvious high-borrowing characteristics. No users borrowed books in Transportation or Agricultural Sciences.

To present the correlation results intuitively, a heat map of Spearman correlation coefficients for borrowing behavior was drawn (see Figure 3). The results show that all coefficients are positive, indicating strong correlations in borrowing preferences for major CLC categories among users of different identities, genders, and educational levels: female undergraduates, female master' s students, and female doctoral students show high correlations with one another, and the categories of books they borrow tend to be more similar than those of users with other educational levels and genders. The correlation between female faculty/staff and female students decreases markedly. Differences in borrowing preferences among male students are relatively pronounced, while male undergraduates show stronger associations with female master' s and doctoral students. At the master' s stage, male and female students have relatively high similarity in borrowing. Overall, within the student group, male and female users show relatively high associations in borrowing characteristics across different educational levels and genders, whereas the borrowing correlations between faculty/staff and students are relatively low.

Figure 3. Heat map of Spearman correlation coefficients for users' borrowing behavior

3.3 Visualization Display of User Portraits

3.3.1 Visualization of Individual User Portraits In this study, the attribute information of an individual user portrait includes name, gender, identity, department, personal avatar, personal profile, similar users, and their proportions. Through the CO-STAR framework, relevant prompts are generated for the name, gender, identity, department, personal avatar, and profile, while similar users and their proportions are obtained through statistical methods.

Based on the above, a large language model is used to generate user portraits and construct AI dynamic user portraits. Figure 4 presents an example of an AI dynamic user portrait at the individual level, mainly showing the following modules: user attributes and descri-

Figure 4. Example interface of an AI user dynamic portrait

···description tags. After the basic information is filtered and aggregated, the user avatar and personal profile generated by the large language model are displayed in the left-side area. User borrowing-preference tags. User behavior tags. Book-borrowing keyword cloud. Profile feedback. Providing users with a channel for evaluating and giving feedback on profile content helps increase user participation, and also enables the profile to be dynamically updated based on user feedback.

3.3.2 Visualization of User Group Profiles

The two most important dimensions of peer relationships are, first, similarity in age and, second, commonality of attributes, reflected in many aspects such as interests and identity. For university students, peers mainly refer to classmates or friends around them[18]. In this study, peers refer to users who belong to the same school or department, have the same identity, and are in the same major but different grades. To present user-group behavioral characteristics and preferences intuitively, a dynamic group profile is constructed. Based on the peer relationship of "school/department-identity-major," user profiles for different groups can be constructed, helping lower-grade students understand the behavioral preferences of higher-grade students and thereby providing a reference for their own academic development.

Figure 5 shows an example interface of an AI dynamic user-group profile generated through large-scale data mining and user behavior analysis. The specific contents are as follows: User group attribute information. Group filtering options are set at the top of the interface, supporting the filtering and generation of user group profiles for specific groups. User group borrowing preferences. The center of the interface displays the top three books by borrowing frequency among group users, the corresponding topic tags, borrowing subject-breadth tags, and personalized recommendations. Group user behavior tags. For example, the right-hand part of Figure 5 takes the group of first-year male master' s students majoring in information science as an example, showing the behavior tags of this user group, including the level of activity in library entry and borrowing, as well as the degree of concentration or dispersion of behavior. Book-borrowing topic keywords. The keywords of books borrowed by the user group are displayed in the form of a word cloud.

Figure 5 Example interface of an AI dynamic user-group profile

4 Conclusions and Prospects

Taking university library user profiles as the research topic, this study proposes a user-profile construction method that combines data mining with large language model technology, and verifies the effectiveness of the proposed method through empirical research. The main contributions of the study include: with multidimensionality, refinement, and intelligence as the core, a library user tag system integrating the generative capabilities of large language models is proposed. The traditional RFM model is improved, and an RFCI user behavior model suitable for library scenarios is proposed. Keywords and topics of borrowed books are analyzed; TF-IDF is used to identify keyword tags for borrowed books, and BERTopic is adopted for topic modeling of book content. Information entropy and Spearman correlation are used to analyze the borrowing behavior characteristics of user groups with different identities, genders, and educational levels, and group-correlation tags are established. Large language models are used to generate personalized user avatars and user profiles, effectively addressing the issue of user privacy protection. The study realizes the construction of AI dynamic user profiles, displaying user characteristics in an intuitive manner, and can promptly reflect changes in user behavior and preferences, providing valuable reference and guidance for the construction of user profiles in university libraries.

Because university teachers and students differ considerably in disciplinary needs, information-acquisition habits, and other aspects, future research should expand the sample scope to include groups of teachers and students from different disciplines, grades, and fields, thereby improving the generalizability of the research conclusions. Meanwhile, with the continuous development of AI technology, future work will further improve and optimize user-profile construction methods, conduct user-participation evaluation experiments, and enhance the accuracy of the research conclusions by collecting user feedback on profile results.

References

[1] Jia Xiaoyu. Research on Precision Knowledge Services in University Libraries Based on Multidimensional User Profiles [D]. Jinan: Shandong University, 2023.

[2] Cooper A. The inmates are running the asylum: why high-tech products drive us crazy and how to restore the sanity [M]. Indianapolis: Sams, 2004: 123–129.

[3] Baxter K, Courage C, Caine K. Understanding your users: a practical guide to user research methods[M]. Burlington: Morgan Kaufmann, 2015: 22–62.

[4] Wang Yibo, Zhang Pengyi. Research themes and evolutionary analysis of domestic user profiles based on the LDA model[J]. *Information Research*, 2024(2): 99–105.

[5] Purificato E, Boratto L, De Luca E W. User modeling and user profiling: a comprehensive survey[J/OL]. arXiv, 2024[2026-01-23]. https://arxiv.org/abs/2402.09660.

[6] Zhang Manyue, Liu Jinya, Zhao Kunhua. User profiles and service implications of research-data reuse practices in the field of biology in China[J]. *Journal of Modern Information*, 2024(6): 33–44.

[7] Hu Rui, Yuan Haifeng, Rui Zhong. Intelligent regulation of HVAC based on user profiles[J]. *Modern Electronics Technique*, 2024(1): 134–139.

[8] Godoy D, Amandi A. User profiling in personal information agents: a survey[J]. *The Knowledge Engineering Review*, 2005(4): 329–361.

[9] Liao Yunping, Lu Mingfang, Yang Siluo. Research on user profiles for smart libraries from the perspective of big data[J]. *Journal of the National Library of China*, 2020(3): 73–82.

[10] Hixson T, Parrott J. Using UX Personas to improve library GIS and data services[J/OL]. *Weave: Journal of Library User Experience*, 2021(2)[2026-03-01]. https://journals.publishing.umich.edu/weaveux/article/id/221/.

[11] Sundt A, Davis E. User personas as a shared lens for library UX[J/OL]. *Weave: Journal of Library User Experience*, 2017(6)[2026-03-01]. https://journals.publishing.umich.edu/weaveux/article/pubid/12535642.0001.601/.

[12] Rohmiyati Y, Tengku Wook T S M, Sahari N, et al. Designing personas for e-resources users in the university libraries[J]. *Computers*, 2023(3): 48.

[13] Dai Ying. Research on the construction of a library user-profile service system[J]. *Journal of the Library Science Society of Sichuan*, 2020(2): 44–48.

[14] Liu Hai' ou, Li Kai, He Xutao, et al. Recommendation of diversified user-profile tags for information cocoons[J]. *Library*, 2022(3): 83–89.

[15] Chen Huixiang. Research on precision library services based on user profiles[D]. Nanjing: Nanjing University, 2018.

[16] Tan, Steinbach. *Introduction to Data Mining*[M]. Translated by Fan Ming, et al. Beijing: Posts and Telecom Press, 2006: 201–202.

[17] Jiqizhixin. How I won the GPT-4 prompt engineering competition[EB/OL]. [2025-12-10]. https://www.jiqizhixin.com/articles/2024-05-14-4.

[18] Liu Rujian. Exploration and application of peer encouragement in reading-promotion activities in university libraries: taking the "One Square Meter of Tranquility" college-student shared reading activity at the Library of Beijing Normal University at Zhuhai as an example[J]. *China Modern Educational Equipment*, 2022(3): 160–163.

Research on the Construction of User Profiles Based on Large Language Models: A Case Study of Peking University Library

Wang Yibo1 **Zhang Pengyi**2

1. Peking University Library, Beijing, 100871

2. Department of Information Management of Peking University, Beijing, 100871

Abstract The construction of user profiles is fundamental to analyzing user characteristics and understanding user needs. This study proposes a method based on large language models (LLMs) and data mining techniques to construct and update user profiles in academic libraries. First, a hierarchical tagging system for academic library user profiles is developed. This system consists of 5 first-level indicators (including user attributes and behaviors), 11 second-level indicators (such as personal information and academic background), and 49 third-level indicators (such as name, department, duration of library visits, and borrowing frequency). Second, the study introduces methods for profile construction and updating, as well as an LLM-based approach for building and updating these profiles. Finally, an empirical study using readers from the Peking University Library is conducted to verify the effectiveness of the proposed methods. Based on this, dynamic profiles for individuals and groups based on LLMs are designed, achieving the visual presentation and dynamic updating of user profiles.

Keywords Large language models; Artificial intelligence; Academic libraries; User profiles; Data mining

Note: Figure translations are in progress. See original paper for figures.