The Usages and Expectations of Multilingual Information Access in Chinese Academic Digital Libraries

Dan Wu  
School of Information Management  
Wuhan University, Wuhan, Hubei  
430072, China  
woodan@whu.edu.cn

Nanhui Gu  
School of Information Management  
Wuhan University, Wuhan, Hubei  
430072, China  
gunanhui@qq.com

Daqing He  
School of Information Sciences  
University of Pittsburgh, Pittsburgh  
PA 15260, USA  
dah44@pitt.edu

ABSTRACT
Digital library, because of its resource demanding and other issues to solve, is an important application of multilingual information access (MLIA). However, the requirements of MLIA systems and applications are not typically addressed or assessed in our evaluations of digital libraries. This paper, therefore, aims to study the usages and expectations of MLIA in Chinese academic digital libraries. We conducted two surveys to study MLIA in current Chinese academic digital libraries and to get to know the users’ real requirements for MLIA in Chinese academic digital libraries. The initial results offer thoughts on specific MLIA functions and insights on future digital library design and developments.

Categories and Subject Descriptors
H.3.7 [Digital Libraries]: User Issues

General Terms
Human Factors, Languages, Experimentation

Keywords
Multilingual Information Access, Digital Library, User Survey, Chinese Academic Libraries

1. INTRODUCTION
Multilingual information access (MLIA) studies the storage, access, retrieval and presentation of information in multiple languages. It is critical for the further integration of people and information at the global scale. One important application area of MLIA is digital library. Digital libraries (DL), which hold large scale digitalized resources, play important roles in media-rich life. Multi-media, multi-linguality and multi-culture are the three major characteristics of digital libraries [1]. As an integration of content and technology, digital libraries contain many MLIA related issues, which include multilingual resource management, multilingual portals, multilingual search, multilingual presentation of results, multilingual question and answering, multilingual text mining etc. Therefore, it is not surprise that increasing number of digital library services have realized the importance of MLIA. For example, the European Commission launched the i2010 Digital Libraries Initiative to enable multilingual access to the contents of Europe’s national libraries [2].

However, MLIA systems have not been widely adopted except a few recent developments such as Google Translate. This is due, in part, to lack of demands in the marketplace, but also, in perhaps greater measure, to the special requirements that may be associated with MLIA applications – the requirements that are not typically addressed or assessed in our research evaluations. One such requirement is end-user support in MLIA systems as end users have greater needs for the translation or summarization of retrieved information [3].

Therefore, to offer insights to MLIA functions in future digital library applications, we conducted a series of studies on usages and expectations of MLIA in the current digital libraries, and our research focus has been on collecting the users’ real requirements directly, and on Chinese academic digital libraries.

In the remainder of this paper, we will first review the related work on MLIA in digital library in section 2; then in section 3, we talk about our research questions and two sets of surveys in detail. Then, we will analyze the results of the two experiments and obtain answers to the research objectives in section 4. Finally, we will conclude with discussions and future works in section 5.

2. RELATED WORK
Research activities associated with MLIA in digital libraries can be divided into three aspects. The first one explores the framework of integrating MLIA with digital libraries. Oard [4] pointed out that users seeking information from a digital library could benefit from the ability to query large multilingual collections with a single language, and thesauri can help to address this challenge by facilitating controlled vocabulary search using terms from several languages. Maybury and Griffith [5] described an integrated environment for information analysts to examine very large multilingual collection. Chen [6] gave an overview of multilingual information access in digital library, in which National Palace Digital Museum was used as an example. Liu at al [7] created Arc, an OAI compliant federated digital library, and discussed how Arc can integrate an existing cross-language retrieval component. Pavani [8] studied Maxwell system Digital Library and identified basic functionalities and components for a multilingual digital library.

The second aspect is about research on multilingual information processing technology for constructing multilingual digital libraries. Bian and Chen [9] discussed cross-language information access to multilingual collections on the Internet. Wang et al [10] investigated the feasibility of exploiting the Web as the corpus source to translate unknown query terms for cross-language

The third one studies the usage of multilingual information resources in existing digital libraries or projects. A research team at the University of Maryland [12] designed the International Children’s Digital Library which selects and processes books from different countries, and presents them in multiple languages simultaneously. Berkeley Public Digital Library [15] provides multilingual resources in eight languages, and has the multilingual catalog search and multilingual reference service. Comparing to the rest of the world, Europe pays more attention to the multilingual issues in digital libraries. Among 14 projects containing multilingual collections funded by the European commission under the Fifth Framework Programme [13], ETRDL project provided multilingual interface in six languages and multiple language text processing, and SCHOLNET was an extension of ETRDL with cross-language search functionality. ECHO was a project about film archives in four languages, and it had cross-language search via controlled vocabulary. MUCHMORE project was for CLIR in medical domain. MultiMatch project is a multilingual and multimedia search engine for cultural heritage [14], which has components for both document and query translation. The European digital library, museum and archive is a single access point to Europe’s cultural heritage in multiple languages.

The reviewed literature shows many digital libraries are the products of collaboration from different countries which naturally produce bilingual or multilingual collections, and these digital libraries serve broader or global user communities with users speaking different languages, but many of them do not have multilingual search capabilities. More importantly, few studies have evaluated the multilingual collections of digital library from the user’s perspective. The lack of user studies is surprising considering the increasing interest in digital library projects.

We identified similar situation inside China. The development of digital libraries in China has a shorter history, and few digital libraries have MLIA services. With China becoming more and more open to the world, it is necessary for Chinese researchers to access foreign language resources easily and to disseminate Chinese achievements to the rest of the world. Therefore, this paper aims to investigate the current usages of MLIA in Chinese academic digital libraries, and to elicit users’ real requirements of MLIA in those digital libraries. Our research results will help Chinese digital libraries to develop and provide MLIA services that truly meet the users’ needs.

3. RESEARCH DESIGN

3.1 Research Questions and Methodology

Our goal in this paper is to examine the usages and expectations of multilingual information access in Chinese academic digital libraries. We here have two research questions:

1) What are the current applications and usages of multilingual information access services in Chinese academic digital libraries?

2) What are the users’ real requirements on the multilingual information access services in Chinese academic digital libraries?

We adopt survey as our data collection method to investigate the above research questions. We have conducted two surveys, and the first one sampled several representative Chinese digital libraries, and interviewed the managers of the libraries to obtain detailed information about the current usage of MLIA services. The other survey used a questionnaire to collect what users from different academic disciplines think of MLIA services in digital libraries.

3.2 Survey Design and Data Collection

3.2.1 Digital library survey

The goal of this survey was to investigate the usages of multilingual resources in Chinese digital libraries. We conducted a pilot study to examine the availability of multilingual resources in major public libraries and major academic libraries. We visited two libraries: the Shenzhen City Library which is a big public library, and the library of Wuhan Branch of Chinese Academy of Sciences which is an big academic library. Our investigation found that the amount of the multilingual resources is quite different between these two kinds of libraries (see Table 1).

<table>
<thead>
<tr>
<th>Library of Wuhan Branch of Chinese Academy of Sciences</th>
<th>Shenzhen Library</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>academic library</td>
</tr>
<tr>
<td>% of total</td>
<td>80%</td>
</tr>
<tr>
<td>Lang.</td>
<td>English (96%), Japanese (1%), French (1%), German (1%), Russian (1%)</td>
</tr>
</tbody>
</table>

Table 2. The Six Selected Academic Digital Libraries

<table>
<thead>
<tr>
<th>ID</th>
<th>Digital Library (DL)</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wuhan University Library</td>
<td><a href="http://lib.whu.edu.cn/">http://lib.whu.edu.cn/</a></td>
</tr>
<tr>
<td>2</td>
<td>Huazhong University of Science and Technology Library</td>
<td><a href="http://www.lib.hust.edu.cn/">http://www.lib.hust.edu.cn/</a></td>
</tr>
<tr>
<td>3</td>
<td>Huazhong Normal University Library</td>
<td><a href="http://lib.ccnu.edu.cn/">http://lib.ccnu.edu.cn/</a></td>
</tr>
<tr>
<td>4</td>
<td>Huazhong Agricultural University Library</td>
<td><a href="http://lib.hzau.edu.cn/">http://lib.hzau.edu.cn/</a></td>
</tr>
<tr>
<td>5</td>
<td>Zhongnan University of Economics and Law Library</td>
<td><a href="http://lib.znufe.edu.cn/">http://lib.znufe.edu.cn/</a></td>
</tr>
<tr>
<td>6</td>
<td>Wuhan University of Technology Library</td>
<td><a href="http://lib.whut.edu.cn/">http://lib.whut.edu.cn/</a></td>
</tr>
</tbody>
</table>

Based on this pilot study, we decided to focus on the digital library services in Chinese academic libraries first. We selected the six academic digital libraries in Table 2 based on the following reasons: First, these are top research universities whose library systems would have multilingual information resources. Second, these universities cover different types of disciplines, such as science and technology, economics and law, agriculture, etc. Third, all these universities are located in Wuhan city, which
To investigate the users’ requirements to the multilingual services, we conducted the second survey by sending a questionnaire to potential digital library users in many academic disciplines. This questionnaire has 99 questions including 11 questions for user’s demographic information, and the remaining 88 questions divided into six categories which include user behaviors, MLIA requirement motivations, multilingual information sources, multilingual services, multilingual searches, and multilingual interfaces (See Table 4). All these categories cover possible aspects where multilingual information resources could be used in digital libraries. Except the questions about users demographic information, all the remaining 88 questions use 5 level Likert scales (1 means “totally disagree”; score 2 means “disagree”; score 3 means “not sure”; score 4 means “agree”; and score 5 means “absolutely agree”).

### Table 3. Questions about the Usages of MLIA

<table>
<thead>
<tr>
<th>Category</th>
<th>Question</th>
<th>Question Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Info.</td>
<td>Q1: What percentage of the digital resources is in foreign languages?</td>
<td>close</td>
</tr>
<tr>
<td></td>
<td>Q2: What are the multiple languages of digital resources?</td>
<td>multiple choices</td>
</tr>
<tr>
<td></td>
<td>Q3: What percentage of the budget is spent on foreign language resources?</td>
<td>close</td>
</tr>
<tr>
<td></td>
<td>Q4: Who are the main users of the foreign language resources?</td>
<td>multiple choices</td>
</tr>
<tr>
<td>Usage of Multilingual Resources</td>
<td>Q5: How often are the multilingual digital resources visited?</td>
<td>Likert Scale</td>
</tr>
<tr>
<td></td>
<td>Q6: Are you satisfied with the usages of the multilingual resources?</td>
<td>Likert Scale</td>
</tr>
<tr>
<td></td>
<td>Q7: Do you think that the current multilingual digital resources can meet users’ needs?</td>
<td>Likert Scale</td>
</tr>
<tr>
<td>Multi-lingual Services</td>
<td>Q8: Do you provide training for using the multilingual resources?</td>
<td>single choice</td>
</tr>
<tr>
<td></td>
<td>Q9: Do you provide multilingual document delivery? How many papers per year?</td>
<td>single choice and close</td>
</tr>
<tr>
<td></td>
<td>Q10: Do you provide multilingual search?</td>
<td>single choice</td>
</tr>
<tr>
<td></td>
<td>Q11: What actions have you adopt to improve the usage of the multilingual resources?</td>
<td>multiple choices</td>
</tr>
</tbody>
</table>

The users we selected are mainly graduate students, teachers, or librarians from the selected universities in Table 2. We totally recruited 78 subjects to fill out the questionnaire. They represent different disciplines, which include information science, library science, computer science, telecommunication, electrical engineering, energy, biology, chemistry, environment, literature, foreign language, publishing, social science, art, etc. To encourage the users to carefully complete the questionnaire, we provide small gifts to them.

### 4. RESULT ANALYSIS

#### 4.1 Current Usages of MLIA in Chinese Academic Digital Libraries

**4.1.1 Basic information**

The results for questions 1-3 in Table 3 are presented in Table 5. The identification number of the digital libraries presented in Table 5 is the same as that in Table 2. The results show that the percentage of the digital resources of foreign languages is pretty high in most of the DLs. Except the library with ID 5, all other DLs have more than 55% of foreign language resources. This means that multilingual digital resources are very important and abundant in those DLs. However, the coverage of languages is not wide. Only three DLs provide multilingual resources beyond English, the remaining three only have English collections. And to the percentages of the budget that the DLs spent on multilingual resources, only 3 DLs told us the amount. However, all these three DLs spend over 55% of their budget on foreign language resources. This means that if those multilingual resources are not effectively used, it is a waste of money.

### Table 5. Result of Basic Information

<table>
<thead>
<tr>
<th>DL ID as in Table 3</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65%</td>
<td>7 languages</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>70%</td>
<td>English only</td>
<td>70%</td>
</tr>
<tr>
<td>3</td>
<td>60%</td>
<td>English only</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>70%</td>
<td>English only</td>
<td>80%</td>
</tr>
<tr>
<td>5</td>
<td>34%</td>
<td>5 languages</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>58.62%</td>
<td>5 languages</td>
<td>55.71%</td>
</tr>
</tbody>
</table>
The multiple choices of question 4 include foreign teacher, foreign student, Chinese teacher, Chinese student, and other staff. Our results show that both foreign and Chinese teachers and students are the main users of these multilingual resources, then are the staffs.

4.1.2 Usage of multilingual resources

For question 5-7 in Table 3, we used Likert Scale to test the usages of multilingual resources in those DLs. 1 to 5 scales represent from “totally not satisfied” to “very satisfied”. As shown in Figure 1, only one DL (DL 4) has a very high visit count of its multilingual resources, all others have only average visit counts. The average satisfaction of the multilingual digital resources is only moderate throughout all DLs. And most managers think that their multilingual resources can generally meet the users’ basic needs, but are not at very satisfied level. These results show that the usages of MLIA in these digital libraries are reasonable but not very satisfying.

![Figure 1. Usages of multilingual resources in the six DLs](image)

4.1.3 Multilingual services

The remaining four questions in Table 3 investigate the multilingual services that those DLs provide. All six DLs have foreign language database training for their users. All of them offer multilingual document delivery services, and the average amount of paper delivered per year is about 4300, with the highest number as 12,000 per year at Wuhan University Library. This demonstrates that multilingual document is required by many users. As for multilingual search, only four of the six DLs have the function. The actions that the six DLs adopt to improve the usage of multilingual resources are as follows: (a) Resources Reorganization (3 DLs); (b) E-journals Navigation (4 DLs); (c) Information Retrieval Training (6 DLs); (d) Multiple Database Search (5 DLs); (e) Personalized Information Customization (4 DLs); (f) Online Translation Tool (None); (g) Cross-language Information Retrieval (1 DL in part of its database); (h) Others including promotion. This shows that the improvements are at general monolingual search and navigation, but the translation related multilingual search services like machine translation and cross-language information retrieval have not been widely adopted.

We can see from the above questions that most digital libraries can provide some basic multilingual services like multilingual portal, multilingual document delivery, multilingual resource training, etc. However, more advanced services, especially services like multilingual information access are missing. We think that this is the future emphasis that digital libraries should pay attention to.

4.2 User Expectations of MLIA in Chinese Academic Digital Libraries

To learn the users’ requirements, we divided our questions into six categories as shown in Table 4. Here we will report the reliability, validity, average score and other results of each category. The statistic tool we used was SPSS 16.0. The measure we used to analyze the internal consistency reliability coefficient was Cronbach’s Alpha, and the measure for validity was KMO (Kaiser-Meyer-Olkin) and Bartlett’s Test. The values of the whole questionnaire and each category are shown in Table 6.

It can be seen that the alpha value of the whole questionnaire is 0.924, which means that the questionnaire is very reliable. As to individual category, only categories 1 and 5 have alpha values between 0.35 and 0.7, which means that the reliability is moderate. All other four categories have alpha values higher than 0.7. For validity, all the six categories have KMO values above 0.5, which demonstrate that the questionnaire is valid. Because of the space, we only present selected findings in the remaining of this section.

| Table 6. Reliability and Validity of the Questionnaire |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| Category | Cronbach’s Alpha | KMO | Bartlett’s Test of Sphericity |
|-----------|-----------------|-----|-----------------|------------------|
| Whole     | 0.924           | -   | -               | -                |
| Category1 | 0.470           | 0.588 | 339.452 | 105  | 0.000 |
| Category2 | 0.771           | 0.674 | 411.538 | 105  | 0.000 |
| Category3 | 0.761           | 0.671 | 689.440 | 231  | 0.000 |
| Category4 | 0.915           | 0.837 | 803.737 | 136  | 0.000 |
| Category5 | 0.662           | 0.725 | 70.748  | 10   | 0.000 |
| Category6 | 0.902           | 0.841 | 602.553 | 91   | 0.000 |

4.2.1 User background

In this category, the character distributions of the 78 subjects are:

1) 56% of the subjects are male, 54% are female;
2) 2.5% are under 20 year old, 73.4% are between 21 to 30, 20.3% are between 31 to 40, 3.8% are over 40;
3) 36.7% received bachelor degree, 48.1% had master degree, 15.2% had doctor degree;
4) 26.6% are junior or senior professors, others are students;
5) 11.4% are from literature, art or history discipline; 32.9% are from social science; 17.7% are from science and technology; and 38% are from engineering;
6) Only 9% of them mastered two foreign languages, and others could only speak English as the second language;
7) Almost all of them had used online database, digital library, search engine, online public access catalogue (OPAC), and online translation tool before. The most frequently used tools were:
   - **online databases**: CNKI and Wanfang Data;
   - **digital libraries or OPACs**: Chaoxing DL, National DL, and university library catalogue;
   - **search engines**: Baidu and Google
4.2.2 User behaviors
Questions Q1-Q15 investigate users’ MLIA behaviors. The average score using 5 level Likert scales of this category is 3.13, and major interesting results are:

1) Users are not sure that they know DL very well.
2) Between multilingual books and journals in traditional paper based media and that in digital format, users are more willing to read multilingual digital resources.
3) Users are not sure that they have difficulty in searching or reading multilingual information.
4) Users are satisfied with their multilingual searches in the languages that they can understand.
5) Users have difficulty in searching for multilingual information in languages that they cannot understand, and they feel that they have the needs to access those information.
6) Users rely on translation tools to help them looking for information in the language that they cannot understand, but they are not satisfied with the translation quality.
7) When users received multilingual information in the languages that they cannot understand, they probably would give it up rather than asking help from friends or librarians.

These findings show that users are eager to and have needs to access to multilingual information, but they have problems in searching and receiving those information that they cannot understand. Translation tools are very helpful here and their quality is not satisfactory.

4.2.3 MLIA requirement motivations
Questions Q16-Q30 investigate users’ motivations for accessing multilingual information. Among all reasons we listed, the average score is 2.92. Below are the reasons that received above 3.0 average scores. From highest to the lowest, they are:

1) I want to know the latest developments in other countries. (score: 3.74)
2) I need to write a literature review. (score: 3.64)
3) I need to finish an assignment. (score: 3.51)
4) I want to conduct a research that is novel. (score: 3.32)
5) I need to conduct my daily work. (score: 3.19)
6) I am asked by some friends for help. (score: 14)

The results show that the main reasons that motivate users to access multilingual information are related to research work rather than daily life. Therefore, we can conclude that MLIA in digital library is more necessary and important than in search engine.

4.2.4 Multilingual information sources
Questions Q31-Q52 examine how users use and evaluate multilingual information sources. The average score of this category is 3.61. The interesting finds are:

1) The sources from which users obtain academic information are search engines first, then digital libraries, then traditional libraries.
2) Users have used multilingual information when they search on search engines, digital libraries and library OPACs.
3) Users use Chinese digital libraries and search engines when they search for Chinese academic information, and they would use tools in other languages when they search for information in other languages.
4) Most users have tried Google Translate cross-language search engine before, and generally think that it is good. But few of them have tried Yahoo Babel Fish cross-language search engine.
5) Users are not satisfied with the multilingual information that current digital libraries, search engines, and traditional libraries provide.
6) Users are not sure whether they are satisfied with the ways that they access to multilingual information.
7) Users have a high expectation of a multilingual integrated digital library that has multilingual information access capabilities for languages that they are not familiar with.

These results show that when users need to find multilingual academic information, they mainly rely on search engines and digital libraries. But they are not satisfied with current multilingual tools, and know little about cross-language search engines. On the other hand, users all hope to use a multilingual integrated DL to access to information.

4.2.5 Multilingual services
Questions Q53-Q69 examine the multilingual services that digital libraries should offer. The average score of this category is 3.98. We identified a set of services in the questionnaire and below are the top services that have the average scores above 4.0:

1) provide clustered multilingual information based on subject or discipline (score: 4.29)
2) provide professional term translation (score: 4.27)
3) provide translation assistance for any language (score: 4.18)
4) provide abstract translation for any language (score: 4.18)
5) provide multilingual expert clustering (score: 4.17)
6) allow users to add term translations (score: 4.12)
7) allow users to customize multilingual RSS feeds (score: 4.10)
8) allow users to correct the wrong translation (score: 4.09)
9) allow users to add tag to multilingual resources (score: 4.06)
10) provide language statistics (score: 4.05)
11) provide multilingual search (score: 4.01)

We can see that users really want some specific MLIA services in their DLs. What they want the most are term translation and abstract translation for any language rather than full-text translation. And they also need clustering functions to organize the multilingual resources as well as some interactive services.

4.2.6 Multilingual search
Questions Q70-Q74 reveal the multilingual search functions that DL should offer. The average score of this category is 3.90. Below are the functions whose scores are higher than 4.0:

1) allow users to search in their native language on multiple language documents (score: 4.29)
2) provide translation assistance for users to choose correct translation of the terms (score: 4.22)
3) translate the abstract of the retrieved documents to the users’ native language (score: 4.05)

Therefore, cross-language information retrieval is necessary for the users. The users also want to translate the retrieved results back to their native language. They need assistance in term translation for translation disambiguation.

4.2.7 Multilingual interface
Questions Q75-Q88 ask for the ideal multilingual interface for DLs. The average score of this category is 3.89. Here are the
features received the average score above 4.0. From the highest to the lowest, they are:

1) DL should classify the retrieved multilingual results based on language. (score: 4.12)
2) For those non-popular language information, DL could use English to describe. (score: 4.08)
3) For term or sentence translation, DL should provide concept explanations for helping to select correct translation. (score: 4.05)
4) For term or sentence translation, DL should provide translation probability for helping to select correct translation. (score: 4.04)
5) DL should offer multilingual OPAC. (score: 4.01)

These findings show that users prefer the DL interface that can provide more assistance for translation.

5. CONCLUSION

In this paper, we conducted two surveys to investigate what are the usages and expectations of multilingual information access in digital library. The results provide answers to the two research questions we proposed:

1) Multilingual resources which cost lots of money to construct are very important and abundant in the digital libraries that we studied, and most of these digital libraries can provide some basic multilingual services. However, the usage of MLIA in these digital libraries has not been very satisfactory, and more advanced services such as translation and multilingual information access are critically needed.

2) The DL users are eager to access to multilingual information, but they have problem in finding the information they cannot understand. Therefore, they need helpful translation tools and multilingual DLs which integrate tools that users frequently use to access to multilingual information. Furthermore, users prefer to use interactive DL interfaces that can provide more MLIA interactions, such as term and abstract translation functions for any language, clustering functions, translation assistance, and translating the retrieved results back to their native language.

Our future work includes: 1) We will confirm our findings through studying more Chinese academic digital libraries. 2) We will study further the difference of the usages and expectations of MLIA in Chinese digital libraries and that of digital libraries in other countries. 3) We will propose methods for better design of Chinese digital libraries, especially at the MLIA services. 4) We will compare the two survey results and articulate how the findings of our study can help the library managers to improve their multilingual services.

6. ACKNOWLEDGEMENT

This work is partially supported by National Social Science Foundation of China under the agreement 09CTQ026, Social Science Foundation of Chinese Education Ministry under the agreement 09YJC870022, and Youth Social Science Foundation of Wuhan University under the agreement 08QNXM39.

7. REFERENCES