Humanities Graduate Students' Use Behavior on Full-Text Databases for Ancient Chinese Books

Ming-der Wu and Shih-chuan Chen

Department of Library & Information Science,
National Taiwan University, Taipei, Taiwan
{mdwu,r92126002}@ntu.edu.tw

Abstract. Digitizing ancient books, especially those related to the humanities, is practiced in many countries. The number of full-text databases in the humanities is increasing. Studies have shown that ancient books are important resources for humanities scholars and researchers. However, comparatively little research has been done concerning the use of those databases. Thirty graduate students majoring in Chinese Literature or History were interviewed in this study. This study attempts to answer the following questions: How do interviewees use the databases? Do they encounter any problems? What do they have to say concerning ancient books in digital or paper form? The results show that humanities graduate students use ancient books databases to locate information concerning their research interests. Most of them are satisfied with the search functions and feel that the databases are convenient to use. However, they comment that the coverage, quality, and search interface could be improved upon. As well, a few graduate students suggest that links to related resources should be added. They state that they do not totally rely upon the databases and continue to use paper sources.

Keywords: Full-text databases, Ancient Chinese books, Humanities graduate students, User behavior.

1 Introduction

A great deal of research has revealed that different forms of information resources hold different degrees of significance for researchers in different fields. In the field of science and technology, journals are more important than books, but in the humanities the reverse is true. Moreover, older books and documents are among the most important resources for research in the humanities. Due to developments in information technology, many resources are now available in the digital form. Some resources' original forms are digital to begin with. Additionally, there are many digitization projects being implemented around the world to transform paper resources into digital ones. Digitization of ancient books concerning the humanities is especially valued by those in academia.

Although the amount of the digital resources is increasing, there has been little research done concerning humanities researchers' use behavior of those resources. Graduate students are much more likely than humanities scholars to use digital resources. Barrett [1]

states that "graduate students are passing through some extremely formative years in their academic careers, where evolving research habits will influence a lifetime of scholarship to follow." Therefore, it is important to understand how graduate students use digital resources. Even though full-text ancient books databases have advantages such as being accessible remotely, time-saving, and having convenient search functions, how graduate students use these resources remains unknown. How do they search the databases? Do they encounter any problems? What do they have to say concerning ancient books in digital or paper form? These issues are worth taking a closer look at.

2   Literature Review

Stone [14] describes the diversity of research materials that humanities scholars require, most frequently in book and journal formats. She mentions that books play a greater role than journals. Talja [15] points out that humanities scholars typically use old or established theories to make sense of new topics. Reynolds [13] depicts that "humanists use books and older materials, rely heavily on their own personal collections, use a variety of material rather than a well-defined core of material."

The digitization of ancient humanities documents is regarded as important in academic circles. For example, looking abroad, a wide variety of Western classics figure heavily in Project Gutenberg [12]. The University of Virginia's Electronic Text Center has collected circa 70,000 digitalized humanities texts, including works by significant authors in the UK and the US, classical British and American novels, Bibles, and much more. Texts in 15 languages, which include *300 Tang Poems*, *Shi Jing*, *The Dream of the Red Chamber*, and many more Chinese classics, are available on the Center's website [5].

Digitization of ancient documents is also valued in Taiwan. The Academia Sinica's Electronic Chinese Ancient Document Project includes the *Twenty-five Dynastic Historie*s, the *Thirteen Confucian Classics*, *Taisho Tripitaka,* Taiwanese historical materials, and other ancient books and records [7]. The National Central Library made efforts to digitize rare books and create a database that is searchable by metadata and offers links to the images of the documents [10]. In addition, the National Palace Museum and other institutions have projects to digitize ancient documents [11].

The number of humanities databases is increasing. However, relatively few studies have been done concerning humanities researchers' use of databases or online information retrieval systems. Some studies were done when information technology was in its infancy; the information technology of that time naturally cannot be compared to today's. Katzen [8] explains why humanities users rarely conducted online searches and indicates that there are relatively few humanities databases available and that retrospective resources are not adequately covered. Watson-Boone [17] indicates that if online databases were easier to use and if humanities scholars were to receive training and change their research habits, they could make better use of existing tools and services. Hoogcarspel [6] states that electronic texts in the humanities have problems such as a lack of information that is vital to humanities scholars such as title screens and a reference to the original print version. Bates [2] and her colleagues published a series of articles sponsored by Getty Information Institute. The two-year study concludes that most humanities scholars make little use

of the online database. Scholars appreciate that the database covers many topics, but complain about the difficulty of its search language and the lack of availability of desired resources. It is interesting to note that scholars regard themselves as experts in their subjects and did not expect to learn anything new from the database. Massey-Burzio [9] investigates to what extent humanities faculty value information technology and view its relation to their research and teaching. She concludes that they "definitely feel the pressure to use and deal with technology." Humanities faculty appreciate the advantages of computer searches, but they feel it is uncomfortable and inconvenient to read off a computer screen for a long time. In a study comparing scholars in different disciplines use of e-journals and databases, Talja and Maula [16] classify humanities scholars as "low level users."

Chen [3] investigates students majoring in Chinese language who had used the ancient Chinese books' full-text databases. The study shows that students had a positive response to the Web version of the database concerning its screen prompts, search functions, response time, search results, results display, and the system when taken as a whole, but were less than impressed with the online help and error messages. In other words, users praised the convenience in using the databases, yet felt frustrated that the system did not provide immediate assistance when they encountered problems during the search process. Wu, Huang and Chen [18] survey the use of full-text databases for ancient Chinese books among humanities faculty. Their results show that professors affirmed the convenience of the databases, but pointed out several issues they encountered while using the system, such as the version of the paper source, accuracy, special characters, layout design, and search and browse functions. It is also noted that there is room for improvement in the coverage, quality of content, and search interface. During the period the research was conducted, all professors used full-text databases; however, they verified their findings with paper versions. Some professors felt that the ancient books databases were helpful and convenient in doing academic research but expressed concern with regard to their graduate students' over-dependency on digital resources.

Graduate students are more apt to use databases than humanities scholars. The study by Delgadillo and Lynch [4] shows that graduate students majoring in history are better than their professors at finding information on the Internet. Yet students' research methods and habits mirrored those of their professors. They conclude that future historians will "still have to adhere to principles of interpretation and creativity" which remain unchanged by information technology. Barrett [1] indicates that humanities graduate students frequently use a variety of electronic resources. Most students learn about electronic resources through their supervisors and colleagues or discover the resources on their own. Students praise several advantages of the electronic resources, including efficiency, the saving of time, the speed of word processing, and the convenience of remotely accessing full-text journals.

3   Methods

Thirty graduate students from the Department of Chinese Literature (15 students) and Department of History (15 students) of National Taiwan University were interviewed to investigate their use behavior of full-text databases for ancient Chinese books.

Participants consisted of 18 male and 12 female students. Selection of participants was based on recommendation by professors in these two departments. Most participants are either working on their master's theses or doctoral dissertations or preparing their research topics. The interview schedule was similar to the one being used in an earlier research conducted by the authors [18] with minor revisions. The questions in the schedule were divided into three categories. The first category was designed to analyze the participants' use behavior on the full-text databases. The second category investigated participants' views on paper versions and databases. The third category explored participants' suggestions for improving the databases. All participants have prior experience in using the databases. In order to help participants better recall their experience of using those databases, at the beginning of each interview, the list of the University Library's full-text databases for ancient Chinese books was presented to each participant. Interviews were recorded and averaged 45 minutes in length.

4   Results

4.1   Use of Full-Text Databases for Ancient Chinese Books

4.1.1   Learning About the Databases
All except one participant learned about the databases from their peers and instructors. Some instructors who commonly used the databases introduced their students to them or taught what to pay special attention to while using them. Six participants mentioned that they also learned of the databases during courses in library instruction.

Most databases provide user manuals or help messages to give information about the usage of the databases. Four participants feel that these manuals or help messages are either too long or too complex and therefore rarely use them. Participants explained that after using the databases a few times they would master it themselves. One participant mentioned that she once asked for a librarian's assistance and learned how to use the databases thereafter.

4.1.2   Installation of Software
A few databases for ancient Chinese books come in an Internet-based CD-ROM version. Users need to borrow the CD-ROM from the library and install the software onto their own computers in order to use the databases. Participants in this study seldom install the software as they feel it to be troublesome, and it is seen as more convenient to use Web-based databases.

In order to display some variant form of Chinese characters, old characters, or rarely-used characters on the computer screen, a few databases provide character creation software for users to install. Without the software, random codes or spaces will appear on the screen instead of the appropriate characters. Six participants mentioned that they do not install the software because they are incapable of doing so.

4.1.3   Methods of Retrieval
While accessing the databases, participants use more search functions than browse functions. Search functions are utilized when participants have keywords (such as

personal names, place names, proper nouns) on hand, whereas browse functions are utilized when participants know or roughly know the location of the sources they are looking for. If participants do not have access to a paper version and need to read the texts quickly, they would browse the full text on the screen.

When conducting a search, most participants used basic searches rather than combined searches (i.e. Boolean logic search). In the basic search, participants mainly used two or more characters as opposed to a single character. They explained that using a single character would retrieve too many results, making it difficult to find relevant items. On the other hand, a few participants preferred to use a single character and then to filter through the long list of results. Participants did not make use of combined searches because they were either unaware that it could be used to perform searches within the databases or because they felt that this kind of search would retrieve fewer results. Although most of the participants used a basic search method only, twenty-one of them responded that they were satisfied with search results done in these databases.

4.1.4   Reading Methods
Twenty-four participants said that they usually read full-text ancient books directly on the screen. Only a few printed out the documents to read them. Some participants mentioned that they copy sentences or paragraphs and save them to their computer for reading at a later time. Those who read directly on the screen mainly did so because their purpose was to locate information. After locating what they were looking for, they would turn to the paper version

4.1.5   Difficulties Encountered
Ancient Chinese books are printed vertically while texts are displayed horizontally in the databases. The horizontal layout was not favored by humanities graduate students because they are not accustomed to reading that way. Five participants mentioned that the divide between the main body of text and its annotations was not clear in certain databases, which make them difficult to read. One participant mentioned that it was difficult to read the annotation because its font was smaller than the main text. Some databases do not state the source from which they were digitized, which made it troublesome for students to determine the appropriate counterpart paper version. Three participants said that they disliked the search interface of some databases and felt that to search in paper versions was easier. "I don't know how to start my search and it is slow to flip the pages on the screen," said one participant. Some participants complained that they needed to click back several pages if they wanted to start a new search. The complexity of the multi-layer structure forced users to click through several pages to locate what they were looking for. If the speed of their home computer's Internet connection was slow, then much time was wasted in using the databases.

4.2   Comments on Paper Versions and Databases

4.2.1   Choice Between Paper Version and Database
When searching for information, most participants would first use the databases to locate the source of the information and then confirm their findings using paper

versions. Participants agreed that searching the databases was comparatively faster and more convenient than searching paper versions. They found that they could gather more results which may otherwise have been missed had paper versions been searched. Also, there is no time and location limit in using the databases, which makes it a better choice. Texts in some ancient books in the databases are punctuated. Four participants said this made the texts much more readable than those in paper versions that did not contain punctuation.

However, participants mentioned that errors showed up in the ancient Chinese books databases, such as incorrect characters and punctuation. They found the accuracy of the databases to be questionable. Therefore, when writing term papers, theses, dissertations, or other important articles, all participants mentioned that they would verify quotations in corresponding paper versions. Most of the time, they used the full-text databases as a tool to find information, but they preferred to use paper versions for further reading.

Participants complained that texts were displayed fragmentally on the screen, meaning that the contextual structure was lost. They preferred to read the paper version to avoid misunderstanding the texts. Moreover, it was found to take more energy to read text on-screen. Participants explained that it was uncomfortable to read text on-screen for long periods of time.

4.2.2 Effect on Theses or Dissertations

Participants responded that full-text databases for ancient books had a noticeable effect on their theses or dissertations, especially with regard to the amount of information they could collect. Participants mentioned that the time spent collecting information would be much longer if the databases did not exist. Thirteen participants   generally felt that using databases broadens the scope of their theses or dissertations. Many of them mentioned that copying and pasting sentences or paragraphs from the full-text databases saves time when they write their theses or dissertations.

4.2.3 Dependence on the Databases

Among the 30 participants, 12 indicated that the existence of full-text databases would not cause them to depend less on paper versions. However, the other 18 participants stated that being able to use the databases has indeed caused them to refer less frequently to paper versions, particularly while in the information gathering stage. They explained that when conducting research, they could not solely depend on a single source, and that the full-text databases helped them gain access to a number of comprehensive resources from related fields. Although some students used paper versions less frequently, this does not mean that the databases can completely replace paper sources. "Paper versions are irreplaceable. They are much comfortable to read and they provide contextual structure," said one participant.

The study by Wu, Huang and Chen [18] indicated that some professors, especially senior professors, are concerned about the effect full-text databases would have on graduate students in their research. Professors worried that graduate students will totally depend on the databases and refer less frequently to paper versions. Most participants remarked that they understood their professors' concern, but they recognized that the full-text database was an increasingly indispensable tool. In the

past, researchers would need to go through the traditional research process, which required extensive reading. Participants said that they spend less time searching for information but still need to build on their ability to comprehend and analyze. They understood that research could still be conducted without the full-text databases but that more time would be needed. Some participants were not concerned, as they felt if their professors had any concerns about the information students gleaned from full-text databases, they could simply request that their students confirm the information in paper versions.

4.3 Suggestions to Full-Text Databases for Ancient Chinese Books

4.3.1 Coverage

Participants felt that the coverage of ancient Chinese books databases should be as exhaustive as possible. Annals, notes, novels, and plays were mentioned as deserving representation in the database in addition to the standard classics. A few participants suggested that important annotations to ancient books should also be covered.

4.3.2 Quality

Participants suggested that errors such as incorrect characters and punctuation should be avoided. They also suggested that images should also be included in the database for the purpose of verifying the errors found in the text file. The databases should give a clear indication of the version of its paper source. Participants questioned the accuracy of the databases and recommended that databases use authorized versions or popular versions in order to improve their credibility. "It is unnecessary for a database to collect all versions, but the version covered should be the best one," said one participant.

4.3.3 Search Interface

Ancient Chinese people usually had several names. A place name may also change over time. Participants suggested that biographical dictionaries, gazetteers or authority files for personal and place names be included in the databases, such that any form of a name used in a search can retrieve all documents containing different evolutions of names. To make a precise search, some participants suggested that the databases should allow users to specify their searches by personal name or place name. A search limited by dates should also be provided. As for browsing, a detailed table of contents would help users to browse any level or part of the book. Participants said databases would be easier to use if there were a unified search interface for all databases. A cross-database search function was also recommended. Generally speaking, participants expected a much more user-friendly search interface.

4.3.4 Display of Search Result

The length of search results is usually short: only the sentence or paragraph that includes the keyword was displayed. Sixteen participants suggested that more sentences or paragraphs should be displayed in order to help the user understand the context.

4.3.5 File Saving

When copying an entire web page to a Word file, participants sometimes encountered problems such as the typeset failing to display correctly or annotations being lost. To

solve such problems, they would need to re-organize the typesetting. They suggest users would benefit if different file formats, such as Word, .txt, or PDF, were provided within the databases.

4.3.6 Other Functions

Participants suggested that hyperlinks to related documents, such as commentaries or interpretations about a given book, be included. This kind of snowball technique links related materials and thus broadens the scope of searches.

Three participants also spoke of including a space for discussion within the databases for users to share their opinions or comments. Personalized settings, such as layout settings or pre-selected databases, could be provided in accordance with individual needs.

5   Conclusions

Researchers in the humanities are stereotypically seen as less likely to access the Internet and use digital resources and to be less familiar with information technology. However, all humanities graduate students interviewed in this study accessed the Internet and ancient books databases. They appreciated the advantages of the databases and have no major problem in searching the databases. Humanities graduate students took a serious approach to utilizing the databases in that they always verified their findings in paper versions.

In comparison with the study made by Wu, Huang and Chen [18], there is no significant difference concerning the use of full-text databases for ancient Chinese books by humanities scholars and graduate students. However, professors relied less on the databases and emphasized the importance of paper versions, while harshly criticizing the flaws found in the databases. Graduate students, on the other hand, were found to be more familiar with information technology. They had more proposed changes, including concrete suggestions concerning the design of the database's search interface. For the most part, professors are familiar with the contents of ancient books. They basically use the database as a tool to locate or cross-check information. Just like the findings of Bates [2], professors do not expect to learn new things from the database. However, graduate students may use the database to discover information to broaden their knowledge on the subjects.

Full-text databases for ancient books are important research resources in the humanities. In the future, the number of databases is certain to rise. Digitization of paper versions is, however, merely the first step. Creators of databases should understand user behavior and utilize the newest information technology to enhance database functions such that databases may become more important and relevant to the humanities.

References

1. Barrett, A.: The Information-seeking Habits of Graduate Student Researchers in the Humanities. Journal of Academic Librarianship 31, 324–331 (2005)
2. Bates, M.J.: The Getty End-User Online Searching Project in the Humanities: Report No. 6: Overview and Conclusions. College and Research Libraries 57(6), 514–523 (1996)

 3. Chen, T.C.: A Case Study on Human-Machine Interaction for the Handy Text Retrieval System. Journal of Library and Information Science 24(2), 65–85 (1998)
 4. Delgadillo, R., Lynch, B.P.: Future Historians: Their Quest for Information. College & Research Libraries 60, 245–259 (1999)
 5. Electronic Text Center: Electronic Text Center Collections, http://etext.lib.virginia.edu/collections/
 6. Hoogcarspel, A.: The Rutgers Inventory of Machine Readable Text in the Humanities: Cataloging and Access. Information Technology and Libraries 13, 27–34 (1994)
 7. Hsieh, C.C., Lin, H.: A Survey of Full-text Data Bases and Related Techniques for Chinese Ancient Documents in Academia Sinica, http://www.sinica.edu.tw/tdbproj/handy1/
 8. Katzen, M.: Application of Computers in the Humanities. Information Processing and Management 22, 259–267 (1986)
 9. Massey-Burzio, V.: The Rush to Technology: A View from the Humanists. Library Trends 47, 620–639 (1999)
10. National Central Library: Guide to Rare Books Image System, http://rarebook.ncl.edu.tw/rbook.cgi/store/frameset.htm
11. National Digital Archives Program, Taiwan: Outcomes of the Program: Rare Books, http://www.ndap.org.tw/2_catalog/visit_folder.php?id=386
12. Project Gutenberg Literary Archive Foundation: Online Book Catalog: Top 100, http://www.gutenberg.org/browse/scores/top
13. Reynolds, J.: A Brave New World: User Studies in the Humanities Enter the Electronic Age. Reference Librarian 49(50), 61–81 (1995)
14. Stone, S.: Humanities Scholars: Information Needs and Uses. Journal of Documentation 38, 292–313 (1982)
15. Talja, S.: Information Sharing in Academic Communities: Types and Levels of Collaboration in Information Seeking and Use. New Review of Information Behavior Research 3, 143–160 (2002)
16. Talja, S., Maula, H.: Reasons for the Use and Non-use of Electronic Journals and Databases: A Domain Analytic Study in Four Scholarly and Disciplines. Journal of Documentation 59, 673–691 (2003)
17. Watson-Boone, R.: The Information Needs and Habits of Humanities Scholars. RQ 34, 203–216 (1994)
18. Wu, M.D., Huang, W.C., Chen, S.C.: Humanities Scholars and Chinese Ancient Books Databases. Journal of Library and Information Studies 4 (2006 in press)