A Chinese Ancient Book Digital Humanities Research Platform to Support Digital Humanities Research

Chih-Ming Chen[1*], Chung Chang[2]

[1] Graduate Institute of Library, Information and Archival Studies, National Chengchi University, Taipei City, Taiwan, ROC; E-mail: chencm@nccu.edu.tw

[2] Graduate Institute of Library, Information and Archival Studies, National Chengchi University, Taipei City, Taiwan, ROC; E-mail: jc09830250@gmail.com

[*] Corresponding author

Purpose—With the rapid development of digital humanities, some digital humanities platforms have been successfully developed to support digital humanities research for humanists. However, most of them have still not provided a friendly digital reading environment and practicable digital humanities analysis tools to support humanists on interpreting texts and exploring characters' social network relationship. Moreover, the advancement of digitization technologies for the retrieval and use of Chinese ancient books is arising an unprecedented challenge and opportunity due to the development of international Chinese study. For these reasons, this paper presents a Chinese ancient books digital humanities research platform (CABDHRP) to support historical Chinese studies. In addition to providing digital archives, digital reading, basic search, and advanced search functions for Chinese ancient books, this platform still provides two novel functions that can more effectively support digital humanities research including an automatic text annotation system (ATAS) for interpreting texts and a character social network relationship map tool (CSNRMT) which can show the character relationships as a knowledge graph with triples containing two character names and their corresponding relationship attribute for exploring characters' social network relationships.

Design/methodology/approach—This study adopted DSpace, an open-source institutional repository system, to serve as a digital archives system for archiving scanned images, metadata, and full texts to develop the CABDHRP for supporting digital humanities research. Moreover, the ATAS developed in the CABDHRP used the Node.js framework to implement the system's front-end and back-end services as well as Application Programming Interface (API) provided by different databases, such as China Biographical Database (CBDB) and TGAZ, was used to retrieve the useful Link Data (LD) sources for interpreting ancient texts. Also, Neo4j which is an open-source graph database management system developed by Neo4j, inc. was used to implement the CSNRMT of the CABDHRP. Finally, JavaScript and jQuery was applied to develop a monitoring program embedded in the CABDHRP to record the use processes from humanists based on xAPI (experience API). To understand the research participants' perception when interpreting the historical texts and character social network relationships with the support of ATAS and CSNRMT, the semi-structured interviews with 21 research participants were preceded.

Findings—An automatic text annotation system (ATAS) embedded in the reading interface of CABDHRP can collect resources from different databases through LD for automatically annotating ancient texts to support digital humanities research. It allows the humanists to refer to resources from diverse databases when interpreting ancient texts as well as provides a friendly text annotation reader for humanists to interpret ancient text through reading. Additionally, the CSNRMT provided by the CABDHRP can semi-automatically identify characters' names based on Chinese word segmentation technology and humanists' support to confirm and analyze characters' social network relationships from Chinese ancient books based on visualizing characters' social networks as a knowledge graph. The CABDHRP not only can stimulate humanists to explore new viewpoints in a

humanistic research, but also can promote the public to emerge the learning interest and awareness of Chinese ancient books.

Originality and value—The proposed novel CABDHRP that contains the advanced features including the automatic word segmentation of Chinese text, automatic Chinese text annotation, semi-automatic character social network analysis, and user behavior analysis to support digital humanities research is different from other existed digital humanities platforms around the world Currently, there is no such kind of digital humanities platform developed for humanists to support digital humanities research.

1. Introduction

Digital humanities (DH) is work at the intersection of digital technology and humanities disciplines (Drucker, 2013). It includes the systematic use of digital contents and tools in the humanities disciplines and makes possible new kinds of teaching and research. In recent years, DH has received a particular attention in both the humanities and computer science fields. Particularly, academic libraries, such as Research Libraries UK and National Central Library of Taiwan, also got started to develop infrastructure to support DH research as well as the close collaboration between librarians and researchers is disclosing new opportunities for librarians to actively collaborate in DH research (Cassella, 2017; Kamposiori, 2017). DH is an interdisciplinary field and is concerned with the intersection of computer science, knowledge management, and a wide range of humanities disciplines (Steiner et al., 2014). Humanities research involves the analysis, comprehension, production, and sharing of a huge amount of digital sources. Moreover, humanists in DH field heavily relied on using digital tools, such as text mining (Widlöcher et al., 2015), annotation (Chen & Tsay, 2017; Sato, Goto, Kimura, & Maeda, 2016), social networks analysis (Uboldi et al., 2013), geographic information system (GIS) (Kallaher & Gamble, 2017), and natural language processing (NLP) technologies (Brooke, Hammond, & Hirst, 2015) to analyze data from digital sources in order to extract useful information, clues, and find new knowledge. However, although a lot of digital tools have already been developed for supporting digital humanities research, many of these tools do not properly fit the needs of the humanities community due to technical difficulties or scientific objectives (Picca & Egloff, 2017). Furthermore, different types of humanists have considerable differences in their knowledge of the collections, requiring varying levels of support, and every individual humanist has their own particular interests and priorities.

In recent years, Chinese ancient books is arising an unprecedented challenge and opportunity due to the rise of international Chinese study around the world. To increase the research efficiency and production, there have been a lot of international Chinese studies to explore the issues about Chinese culture, religion, economy, and politics through Chinese ancient books or literatures with the support of digital humanity tools. For example, Hsiang, Chen et al. (2012) was inspired by Imperial Court documents from the Ming and Qing dynasties because a Memorial often quotes earlier Imperial Edicts as the directives for the actions being reported, while an Imperial Edict may cite earlier Memorials as reasons for issuing the decree. Their study thus explored the citation relation among the Imperial Edicts and the Memorials through the developed digital humanity tool based on Taiwan-related Imperial Court documents from the Ming and Qing dynasties archived in the Taiwan History Digital Library (THDL) (http://thdl.ntu.edu.tw/index.html). Moreover, land deed research played a unique role in modern Chinese history. Land deeds were the only proof of ownership in pre-1900 Taiwan. They are indispensable for the studies of Taiwan's social, anthropological, and economic evolution. Based on a large amount of historical land deeds, Chen et al. (2013) developed a digital humanity tool to discover two important relations－ successive transactions and allotment agreements involving the same property as well as presented a notion of land transitivity graph to capture the transitivity embedded in these land deeds' transactions. Also, Chan et al. (2015) presented a spatiotemporal model based on a geographical information system,

spatial statistics, and the Shannon entropy for elucidating the evolution of market towns spanning and examining the spatial and hierarchical relationships of the market towns from before 1550 to 1980 through a large quantity of historical Chinese ancient books.

To provide more convenient and friendly research environments for humanists engaging in international Chinese study, increasing effort has been made not only to make digital contents and tools available, but also to create a complete virtual research environments (VREs) that provide interpretative frameworks for making sense of cultural artefacts (Steiner et al., 2014). Such VREs support conceptualizing, visualizing and analyzing information, as well as humanists can collaboratively work on it. VREs usually do not consist of one monolithic technology, but cover a collection of tools assembled in one place to assist research tasks and processes (Steiner et al., 2014). This study surveyed several current digital humanities platforms in Taiwan and around the world, including CULTURA system (Steiner et al., 2014), CText (https://ctext.org/zh), Scripta Sinica Database (http://hanchi.ihp.sinica.edu.tw/ihp/hanji.htm), TextGrid (Neuroth et al., 2011), Taiwan History Digital Library (THDL) (http://thdl.ntu.edu.tw/index.html), CBETA Research Platform (CBETA-RP) (http://cbeta-rp.dila.edu.tw/), and MARKUS semi-automatic text annotation system (http://dh.chinese-empires.eu/markus/beta/) and tried to compare their features in order to develop a more practicable digital humanities research platform to support digital humanities research. Among the developed digital humanities platforms, Steiner et al. (2014) presented the CULTURA system and services and the two collections that have been used for testing and deploying the digital humanities research environment. The services of the CULTURA system include personalized search tools, faceted search tools, annotators, social network visualization tools, and recommenders. Additionally, the CText is a Chinese ancient text database, which contains the full text of various Chinese texts of philosophical, historical, or linguistic interest from the pre-Qin era through to the Han Dynasty and beyond, and makes the best possible use of modern information technology to aid humanities research. The Scripta Sinica Database contains almost all of the important Chinese classics, especially those related to Chinese history. This database provides scholars, students, and the general public with an excellent full-text database and search engine for the study of Chinese history and culture. Furthermore, TextGrid is a virtual research environment that provides services for supporting researchers in the arts and humanities and tools for the analysis of text data and supports the curation of research data by means of grid technology (Neuroth et al., 2011). The THDL covers about 80 percent of all primary Chinese historical materials about Taiwan before 1895. The primary functions of THDL for supporting digital humanities research include full-text search, techniques and interfaces for classifying and exploring a query result as a sub-collection, term frequency analysis, and referential tools (Chen et al., 2007). The CBETA-RP provides a friendly online reading interface with complete content and handy digital tools. Besides, full text search, dictionaries look up, person and place references, even the statistics of term, and relevant bibliography are also provided in the CBETA-RP, whereas MARKUS is a famous semi-automatic text annotation system and supports online text reading. However, most of digital humanities platforms currently developed have still not provided complete data analysis functions to support digital humanities research. This study thus presents a Chinese ancient books digital humanities research platform (CABDHRP) which provides basic data analysis functions, such as full-text search with a single word or double words, term frequency analysis and post-query classification, as well as advanced data analysis functions, such as automatic text annotation, textual analysis, character social network analysis, and user behavior analysis, to support digital humanities research.

Table I shows the function comparison of the seven digital humanities platforms. It was discovered that the Scripta Sinica Database presents the richest full-text data collection, the CBETA-RP merely includes Buddhist texts, and MARKUS semi-automatic text annotation system does not show the full-text data collection function. Currently, the CABDHRP merely contains one hundred and ten of Ming Dynasty writings, but the archived contents are rapidly increasing with

time. The CULTURA system consists of multiple distinct services including personalized search tools, faceted search tools, annotators, social network visualization tools, and recommenders (Steiner et al., 2014). Besides CABDHRP, the other six digital humanities platforms are lack of automatic segmentation of Chinese word; the automatic annotation function exists in the CABDHRP, CULTURA, CText, THDL, and MARKUS semi-automatic text annotation system; and, each system presents the search function and external reference function. In terms of external reference, Scripta Sinica Database could connect to Qing Officials Query System, Taiwan History Digital Library presents variant databases, CBETA-RP merely includes Buddhist text databases, and MARKUS semi-automatic text annotation system shows the most external reference resources, containing Wikipedia, CBDB, TGAZ, and ZDict. Merely the CABDHRP, CULTURA, CText, THDL, and CBETA-RP show the function of word frequency statistics; merely THDL presents the function of bookmark notes; and, MARKUS semi-automatic text annotation system has the function of adding annotations to a text. The advanced features of the proposed CABDHRP that contain the automatic word segmentation of Chinese text, automatic Chinese text annotation based on LD, semi-automatic social network analysis of human-machine interaction, and user behavior analysis based on recoding use process of humanists are different from the other six existing digital humanities platforms. Merely CText, CBETA, and MARKUS present the function of open API that can share their digital humanities tools' functions or archived contents with other digital humanities research platforms. These advanced features in the proposed CABDHRP will be detailed in section 3.2.

Table I. Function comparison of different digital humanities platforms

Function comparison of different digital humanities platforms	CABDHRP	CULTURA	CText	Scripta Sinica Database	Taiwan History Digital Library	CBETA Research Platform	MARKUS semi-automatic text annotation system
Full-text data collection	○	○	○	○	○	○	X
Automatic segmentation of Chinese word	○	X	X	X	X	X	X
Automatic annotation	○	○	○	X	○	X	○
Term search	○	○	○	○	○	○	○
External reference	○	○	○	○	○	○	○
Word frequency statistics	○	○	○	X	○	○	X
Bookmark note	X	X	X	○	○	X	X
Social network visualization tool	○	○	○	X	X	X	○
Recommender	X	○	X	X	X	X	X
Recording use process	○	X	X	X	X	X	X
Open API	X	X	○	X	X	○	○

"O" means that a feature is available; "X" means that a feature is not available

2. Literature Review

2.1 Current states, future challenges, and development trends of digital humanities

DH, formally known as humanities computing, is born of the encounter between traditional humanities and computational methods (Burdick et al., 2012). Wang and Inaba (2009) indicated that there is no clear subdiscipline in DH and DH is still expanding its research domain. Tang et al. (2017) also indicated that DH is continuing to expand and become more inclusive as well as its knowledge is integrated. Therón et al. (2016) argued that the current key issues of DH involve the inclusion of citizens in the creation and consumption of the cultural resources offered, the volume and complexity of datasets, and available infrastructures (Therón et al., 2016). Cassella (2017) indicated that a very special relationship is developing between digital humanities and digital libraries in academic libraries when humanists are more and more massive use of the digital library collections to study and to research.

Kaplan (2015) proposed three challenging issues on DH research derived from big data generated by a huge amount of digital contents. The first challenge is how to focus on the processing and interpretations of large cultural datasets, the second challenge is how to concern digital culture at large, and the third challenge is how to deal with the experience of big data. Moreover, many previous studies on DH have focused on the textual contents; however, new work that develops digital tools to analyze the visual layout and content of books and manuscripts based on computer vision techniques is emerging (Rushmeier et al., 2015). Apparently, challenges and increasing research areas identified in digital humanities include how to incorporate the different types of digital resource. In addition, Therón et al. (2016) also indicated that DH field is currently facing the challenge of proposing frameworks and systems that can be standardized for the different disciplines with similar but heterogeneous datasets. Therefore, developing a digital humanities research system that can simultaneously archive digital contents from different disciplines and heterogeneous datasets and perform data analysis by using a set of commonly used and standardized digital tools is definitely needed.

Upadhyay and Upadhyay (2017) claimed that the DH needs to increase the development and demand in the field of computational science, social network theory and analysis, digital literacy, computational pedagogy, computational literacy, machine and deep learning, neuro and cognitive learning and analysis in the future. It is obvious that DH is an emerging and interdisciplinary research field aiming at enhancing and redefining traditional humanities scholarship through digital means.

2.2 The developed digital tools for supporting digital humanities research

DH aims to use a digital-based revolutionary new way to carry out enhanced forms of humanities research more effectively and efficiently. Martin-Rodilla and Gonzalez-Perez (2016) classified the existing works that already developed digital tools to support digital humanities research contain the manual approaches, semi-automatic approaches, and automatic approaches. In the manual approaches, humanists performed a manual data analysis by using digital tools. In the semi-automatic approaches, digital tools offered suggestions based on data analysis results for humanists to further perform data analysis procedures. In the automatic approaches, digital tools automatically performed data analysis based on the computer algorithms for humanists to explore new clues or knowledge.

Currently, in the manual approaches, there have been some existing systems for manually annotating historical documents that provide functions to append information to historical iconographic or text documents to support digital humanities research, such as SMART-GS

(Hashimoto, 2014) and collaborative annotation system (CAS) (Chen & Tsay, 2017). Chen and Tsay (2017) proposed a novel collaborative annotation system (CAS) with four types of multimedia annotations including text annotation, picture annotation, voice annotation, and video annotation which can be embedded into any HTML Web pages to enable humanists to collaboratively add and manage annotations on these pages and provide a shared mechanism for discussing the shared annotations from humanists. The study indicated that the proposed CAS has high potential in the applications of digital humanities research, including discovering the useful knowledge from a large amount of annotations on the historical texts based on data mining technologies, exploring a lot of user interactive data generated in the collaborative annotation process based on social networks analysis, and recording the operation processes of humanities scholars on the CAS to evaluate humanities scholars' research behaviors. Moreover, Sato et al. (2016) also presented a Web-based prototype system for collaboratively making annotations on historical documents by multiple humanities researchers who are distant from each other. The proposed system can support multiple users to make annotations to the same document simultaneously as well as provide suggestion of annotation by using information such as existing annotation strings and their surrounding words in the text. Also, Haslhofer, Robitza and Lagoze (2013) indicated that semantic tagging is a collaborative process in which a user selects and associates Web resources drawn from a knowledge context. Their study applied the semantic tagging technique in the specific context of online historical maps and allowed humanists to annotate and tag them. Compared to the label-based tagging technique, the semantic tagging technique not only does not affect tag type and category distributions, but also leads to higher user satisfaction.

In the semi-automatic approaches, Picca and Egloff (2017) presented a new python library called as The Digital Humanities ToolKit (DHTK) to provide a fast and intuitive tool to exploit the semantic knowledge resources from Linked Open Data (LOD). The key features of the DHTK include digital humanities oriented, ease of use, modularity, efficiency, extensibility, and documentation. Uboldi et al. (2013) designed the Knot, which is a digital tool for exploring historical social networks, to provide humanists with an environment for exploring multi-dimensional and heterogeneous data, allowing them to discover and create explicit and implicit relationships between people, places, and events. Moreover, Jackson (2017) used social network analysis to identify an additional role played by Duncan II Earl of Fife and applied network density model to identify opinion leaders in medieval Scotland based on the People of Medieval Scotland database. Also, Brooke et al. (2015) introduced a software tool, GutenTag, which is aimed at giving literary researchers directly access to natural language processing techniques for the analysis of texts in the Project Gutenberg corpus. Harris et al. (2014) presented the anatomy of Samtla, which is an online integrated research environment designed in collaboration with historians and linguists, to facilitate the study of digitized texts written in any language. Samtla is fundamentally different from standard text search/mining systems which rely on the bag-of-words representation of text. In contrast, Samtla emphasizes the retrieval and discovery of fuzzy text patterns/motifs which are of critical importance to digital humanities research. Furthermore, Luczak-Roesch et al. (2018) presented a novel tool for the digital humanities that leverages temporal data mining, network science, and visual analytics. Their preliminary analytical results show that this approach facilitates a new collaborative methodological practice that is a hybrid of close and distant reading. Pastorelli (2017) applied hierarchical ascendant clustering algorithm to group the manuscripts of the Old Latin witnesses of the Gospel of John into clusters presented with the calculation of distances between readings, then between witnesses to enable the construction of trees illustrating the textual taxonomy obtained.

In the automatic approaches, Yohan et al. (2014) proposed a rule based named entity recognition and classification system which can automatically recognize named entities in text and assign them with the appropriate categories for Telugu language. Additionally, Kestemont and Gussem (2017) applied a layered neural network architecture from the field of deep representation

learning to automatically solve two sequence tagging tasks including part-of-speech tagging and lemmatization. The result of their study shows acceptable performance, which are on par with previously reported studies. Although there have been a lot of digital tools to be successfully developed to support digital humanities research, Harris et al. (2014) argued that there is still a huge gap between what humanists actually want and what digital tools can do in terms of functionality and usability.

3. The Developed CABDHRP

3.1 System architecture

Widlöcher et al. (2015) claimed that developing an integrated environment to support DH research is needed. Their study emphasized that reading interface, information retrieval, and data mining should not be regarded as totally separate tasks when developing an integrated environment to support DH research. Based on the perspective, this study developed a CABDHRP that includes three main parts including the digital reading interface module, basic information retrieval and analysis function module, and advanced text and social network mining module to support DH research. This study adopted DSpace (Chen et al., 2012), an open-source institutional repository system, to implement the CABDHRP embedded with basic and emerging digital tools for supporting DH research. Although DSpace has sufficient functionalities for archiving, searching, and managing different types of digital contents, the predefined metadata schema Dublin Core (DC) needs to be revised to archive digital targets, such as Chinese ancient books, maps, and music. Additionally, the user interface of DSpace for reading digital content cannot satisfy humanities scholars' needs due to very primitive and simple interface. Therefore, the developed CABDHRP based on DSpace has to implement a friendly reader that can simultaneously display the scanned image of Chinese ancient text and the corresponding full-text and integrate many useful reading functions, such as automation word segmentation, automatic text annotation, and character social network analysis, on the reader. These functions are detailed in the later section. Currently, the CABDHRP has archived one hundred and ten of Ming Dynasty writings with scanned images, metadata, and full texts. To support digital humanities research via using Ming Dynasty writings in depth, having a larger body of Ming Dynasty writings will undoubtedly provide benefits in exploring potential research issues to the humanists who are interested in international Chinese study. Figure 1 shows the system architecture of the proposed CABDHRP in this study. The components of the system are explained in details as follows.

(1) Digital reading interface module

The digital reading interface simultaneously contains an automatic text annotation tool and a manual annotation tool, but the manual annotation tool has still not been integrated in the CABDHRP so far. An automatic annotation tool that can collect resources from different databases through LD for automatically annotating Chinese ancient texts was developed in this study to support digital humanities research. It allows the humanists referring to resources from diverse databases when interpreting Chinese ancient texts as well as provides a friendly text annotation reader for humanists interpreting ancient text through reading. Moreover, this study adopted Asynchronous JavaScript (Ajax) and Model-View-Controller (MVC) framework to implement a manual text annotation tool with four types of multimedia annotations including text annotation, picture annotation, voice annotation, and video annotation for knowledge creating, archiving, and sharing services to support digital humanities. The manual text annotation tool has been successfully applied in digital curation, crowdsourcing and digital humanities because of existing strong relations among them (Chen & Tsay, 2017). Specially, in addition to annotating a digital text, this tool also offers photo annotation functions that allow a user to give various types of annotations

on a photo, such as an ancient map.

(2) Basic information retrieval and analysis function module

The basic function module contains the term frequency analysis tool, geographic information analysis (GIS) tool, intelligent search tool, and user behavior recording tool. For textual corpus studied in DH, a search tool is an obvious necessity. The term frequency analysis tool can help humanist to observe the distribution state of a term searched by a humanist in a collection. Moreover, spatial analysis has become one of the primary focus areas for many different types of humanities scholars. The geographic information analysis tool aims to interpret data through spatial relationships. This study has gotten started to apply TGAZ API, which is a read-only interface designed to search the contents of the China Historical GIS placename database, to implement the geographic information analysis tool for supporting the spatial analysis of Chinese ancient books. The intelligent search tool contains full-text search and metadata search with post-query classification according to a humanist's search query. Full-text search aims to help humanist find the documents that contain a specific term, whereas metadata search aims to help humanist find the documents with a specific metadata attribute. Also, since a huge amount of humanists' behaviors of operating the CABDHRP are worthily collected, this study thus developed the user behavior recording tool based on Experience-API (xAPI), which is a new generation learning process record standard developed by Advanced Distributed Learning (ADL) (Tin Can API, 2015). Recording the operation processes of humanities scholars on the CABDHRP aims to evaluate humanities scholars' research behaviors for improving the usability of system functions, not peeking the privacy of system users. "Actor," "verb" and "object" are the record formats generated by the learning process recorded through xAPI. The actor is the executor of behaviors, and could be an individual or a group, for example, a user. The verb is the type of action performed by the actor, for example, clicked or logged in. The object is the entity that interacts with the actor, for example, an image. xAPI provides the flexible design architecture to allow us to self-define the verb names for the recorded learning process.

(3) Advanced text and social network mining module

The advanced text and social network mining module contains the user behavior analysis tool, social network analysis tool, and textual mining tool. The user behavior analysis tool aims at purposively, accurately, and authentically recording humanist's behaviors while operating the CABDHRP for digital humanities research by using the xAPI technology. Such records can include details like log-in usage time, system's operating behavior, text viewing time, and interactions with other humanists. The social network analysis tool aims to help humanists examine the relationships between the characters extracted from the writings of Chinese ancient books via visualizing social networks. This tool provides a knowledge graph, which describes and stores character relationships as triples as well as represents the knowledge graph as a multi-relational graph consisting of characters as nodes and several considered relationship attributes as different types of edges. In other words, a knowledge graph represents some concepts (e.g., people, places, events) and their semantic relationships (Haslhofer et al., 2018; Krause et al., 2016). A knowledge graph could be vehicles for formalizing and connecting findings and insights derived from the analysis of possibly large-scale digital corpora in the digital humanities domain (Haslhofer et al., 2018). The textual mining tool aims to help humanists explore implicit information or knowledge from the Chinese ancient books based on data mining technologies including classification, clustering, association rule, and prediction.
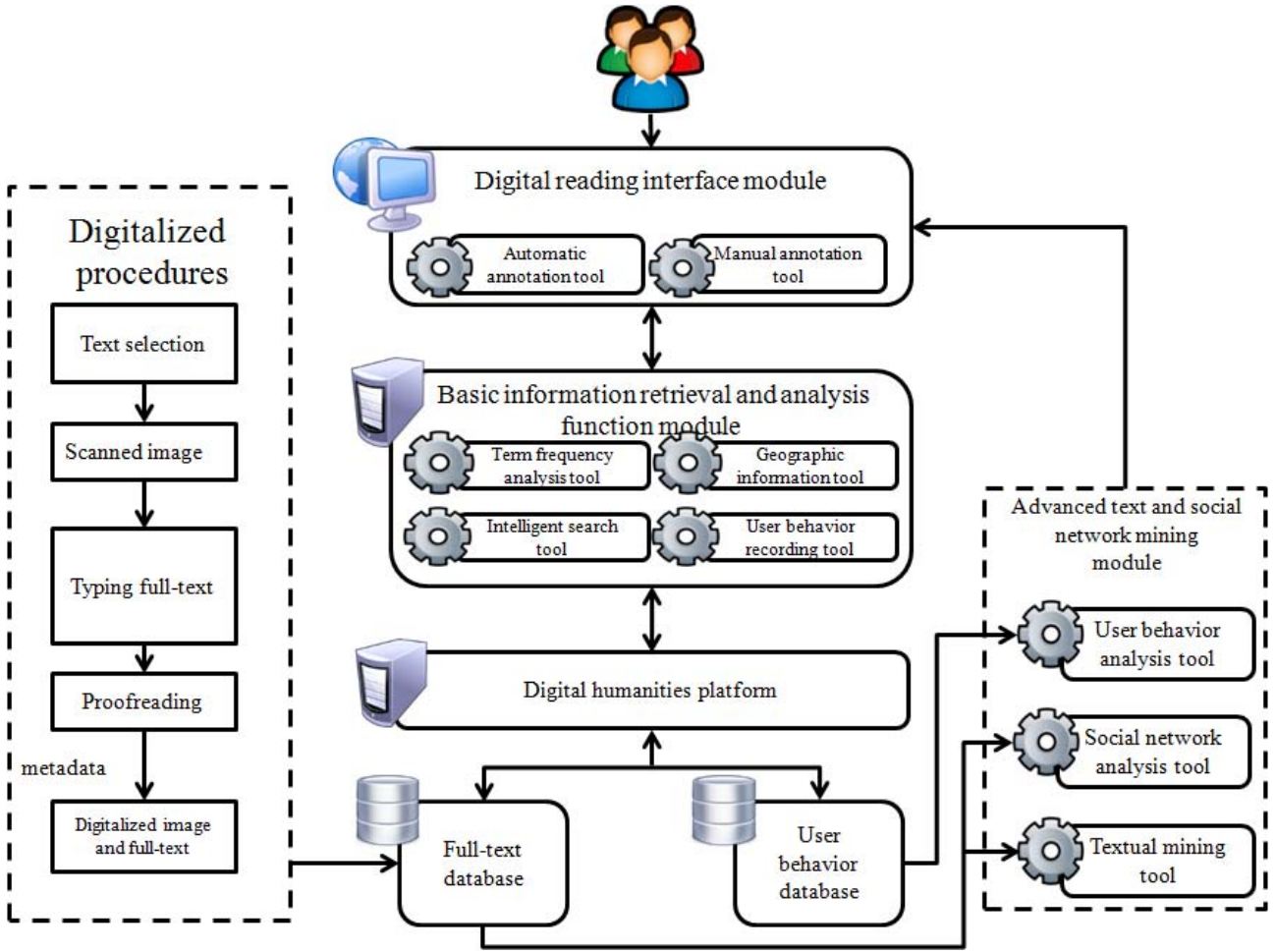
Figure 1. The system architecture of the developed CABDHRP

3.2 System's main functions

This section mainly focuses on introducing the main functions provided by the developed CABDHRP. Figure 2 shows the homepage of the CABDHRP on Ming Dynasty Chinese ancient books. The CABDHRP not only can aid a humanist to explore new facets in the humanistic research, but also can promote the public to emerge the learning interest and awareness of ancient books. Figure 3 shows the full-text search and post-query classification results according to a humanist's search query. Undoubtedly, providing full text search is the easiest way to help humanists find the documents containing a particular item in a large dataset, either in isolation or together with other browsing options, such as faceted search. The post-query classification aims to classify the documents of the resulting set of a query according to several predefined dimensions, such as year and author. Figures 4 and 5 show the comparison charts of term frequency distribution of two terms simultaneously searched by a humanist on year and collection, respectively. This function can help humanist observe the variation of two terms' frequency distribution with year or collection.

Figure 2. The homepage of the CABDHRP on Ming Dynasty Chinese ancient books

Figure 3. The full-text search and post-query classification results of a humanist's search query

Figure 4. The comparison chart of term frequency distribution of two terms searched by a humanist based on year

Figure 5. The comparison chart of term frequency distribution of two terms searched by a humanist based on collection

When a humanist enters the reading interface of the CABDHRP for text reading, the system would segment the entire text into more than two terms based on Jieba Chinese parser which is a lexicon-based Chinese word segmentation system and display these segmented terms with blue

highlight. When a humanist moves the used mouse cursor to the term with blue highlight, the system would change the term with red highlight to remind that the term has available annotation information from the databases with LD for the humanist (Fig. 6). The system would present the annotation content after the humanist clicks on the term with red highlight. On the top of the annotation, the annotations of five default databases, including Wikipedia, CBDB, TGAZ, Moedict, and EC dictionary, could be mutually referred. It would not be displayed when there is no LD in the five default databases for a term. When a humanist clicks on a term with available annotation to view the automatic annotation content, the annotations from different LD resources are displayed with default order in the system. The user could rate whether such annotation is helpfulness or not based on self-judging. The humanist would enhance the annotation score if she/he rated the annotation as a useful database; on the other hand, the annotation score would be reduced if she/he rated the annotation as a helpless one. The system would calculate the accumulated annotation scores of different LD resources. When the humanist re-clicks on the annotation, the annotations from different LD resources would be ordered according to the accumulated annotation scores; ones with higher annotation scores are displayed in higher priority (Fig. 6).
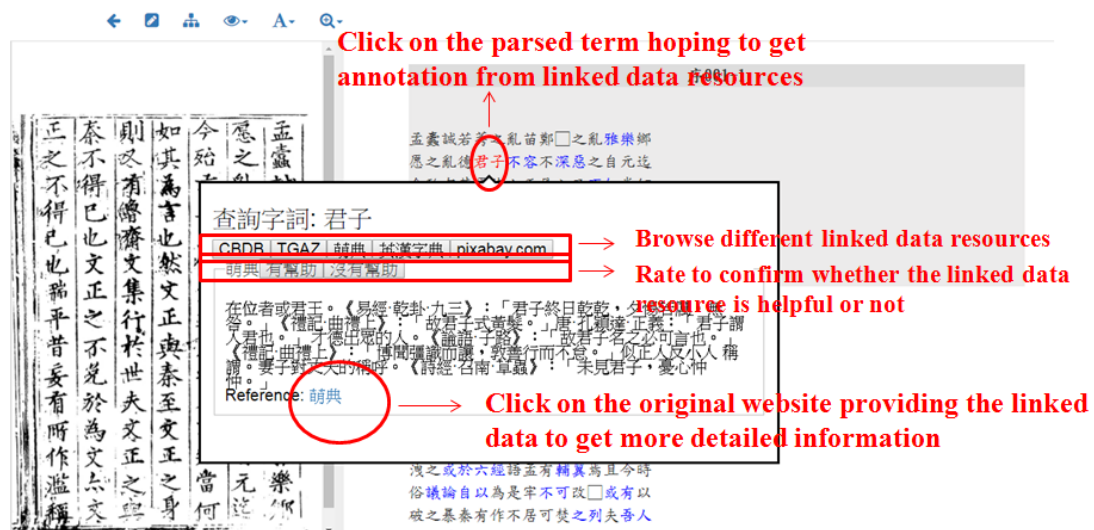

Figure 6. The user interface of reading Chinese ancient text on the CABDHRP

The ATAS would be affected by the correctness of Chinese word segmentation and some unknown terms that cannot be identified by the Chinese word segmentation system. Currently, the accuracy of Chinese word segmentation for the texts of Ming Dynasty ancient books has still not been good enough due to a lot of one-gram words with meaning in Chinese ancient texts and lack of a Chinese word segmentation lexicon that can parse Chinese ancient texts well. Therefore, ATAS would not show automatic annotations based on LD from different databases for the unknown terms appearing in a text. In this case, the humanist could select a term without annotations in the system. When there are annotations for the term, the humanist could consider adding the term as a new term. After clicking on the "Add new term button", the system would pop out a window for confirmation, and then the new term is added to the parser lexicon of the Chinese word segmentation system Jieba so that the correctness of Chinese word segmentation can be gradually promoted due to the human computer interaction. The system would precede correct word segmentation and automatic annotation of the term in the next loading for the reader (Fig. 7).
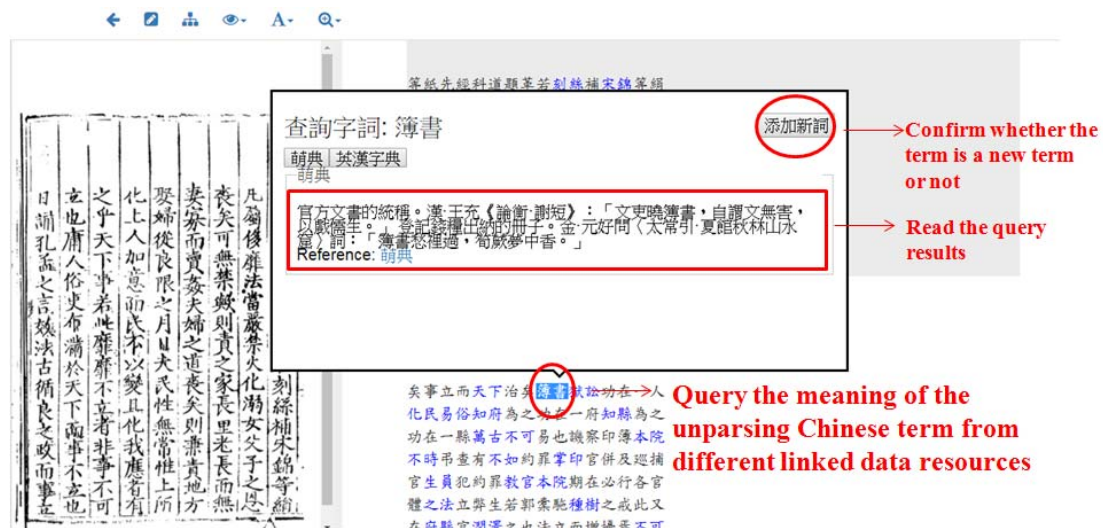
Figure 7. The user interface of adding a new term into the lexicon for enhancing Chinese word segmentation performance on the CABDHRP

Applying social network analysis for DH research aims to examine relationships between individuals, organizations, or groups that interact with each other. These relationships represented by social networks are often visualized to help humanist understand how they are tied together and assist in drawing conclusions and raising further research questions. Thus, the CABDHRP also develops a character social network relationship map tool (CSNRMT) that can semi-automatically assist humanists to more efficiently and accurately explore the character social network relationships from texts for useful research findings through human-computer interaction. The CSNRMT developed on the CABDHRP provides a friendly user interface and functions, allowing humanists interpreting character social network relationships with the support of character social network relationship map, reading interface, text and external search, and notebook provided by the system, in the text interpretation process. The overall system interface of the CSNRMT is shown in Fig. 8 and the functions are explained as below.
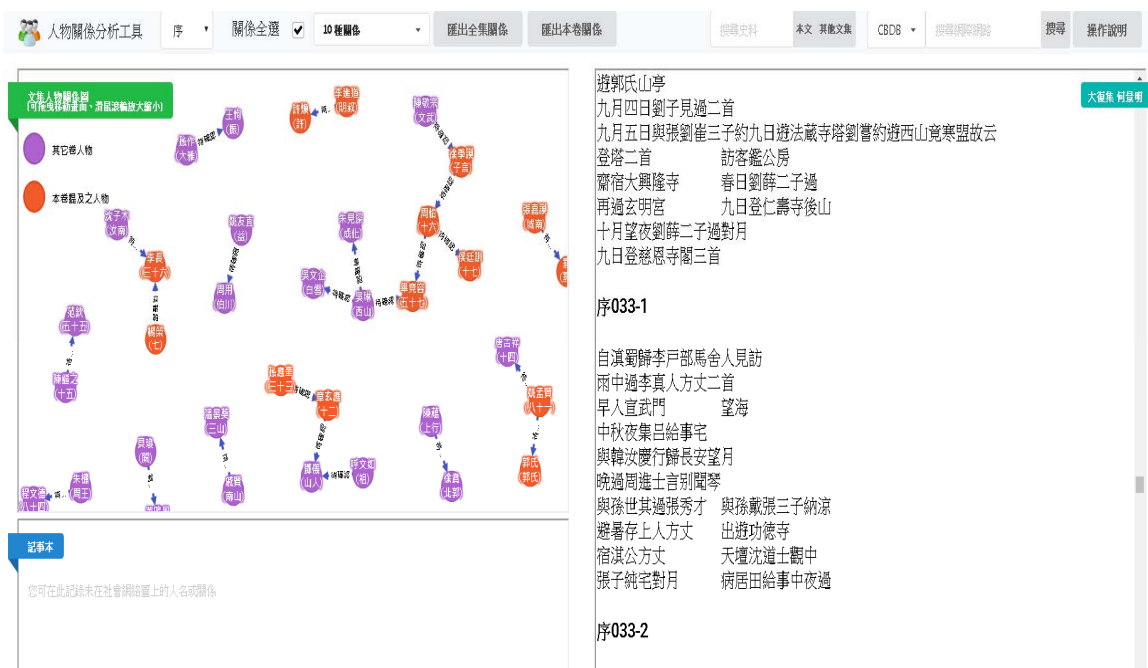

Figure 8. The overall system interface of the CSNRMT

13

1. The user interface of the CSNRMT

    Users could view the character social network relationship in texts by using the user interface of the CSNRMT shown as Fig. 9. The orange and blue nodes in the figure are all characters in the text automatically determined by the system. The orange character nodes are the characters appearing in the text of current reading text paragraph, and the purple character nodes are the characters appearing in other text paragraphs. For example, one of character nodes appearing in the CSNRMT is shown as Fig. 10 and each character node will display two character names and the relationship attribute between the two character names. The system would pre-judge the character's relationship being unknown based on the character relationship analyzer. The edition interface of character social network relationship is shown in Fig. 11. By clicking on the character relationship line in the character social network relationship map, the character social network relationship matrix compiler is started for editing character social network relationship. The edited relationship would be immediately updated on the character social network relationship map.
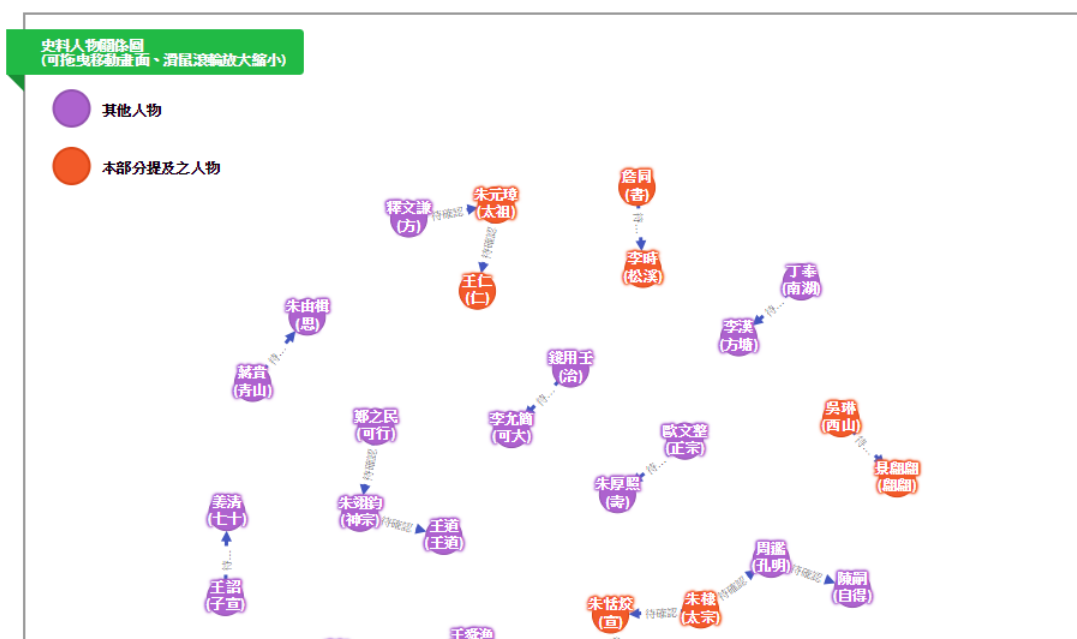

Figure 9. The user interface of the character social network relationship map

Figure 10. The social network relationship between two characters

Figure 11. The edition interface of character social network relationship

2. The user interface of reading a text

The reading interface of a text, as shown in Fig. 12, is integrated with the character social network relationship map and external search function. The character social network relationship map would correspond to the changes according to the text paragraphs in the reading interface. When users click on a character node appearing on the paragraph in the character social network relationship map, the location of the character in the text is highlighted on the reading interface for the users analyzing character social network relationship. Marking a word by using a mouse could link with the external search function for querying extra information about the word. Detailed operation mechanisms for the function are explained in the text search and external search.

Figure 12. The user interface of reading a text

3. The user interface of text search and external search

The text search allows users searching characters or words in a text and will highlight the search results on the text. The external search allows users searching required information through the default external databases for interpreting character social network relationship in a text. It provides external search through various databases including character, online encyclopedia, search engine, and dictionary, such as CBDB, Moedict, Wikipedia, Baidu Baike, Google, Zdic, dictionary of Chinese Character Variants, and Kangxi Dictionary as shown in Fig. 13. A user simply needs to mark a word in a text, and the word will be automatically filled in the external search box for searching.

Figure 13. The user interface of external search

4. The user interface of notebook

The user interface of notebook shown as Fig. 14 allows users recording found information during interpreting a text, expecting that users could reduce the time for switching between windows to complete the interpretation of character social network relationship under a single interface.

Figure 14. The user interface of notebook

5. The user interface of exporting character social network relationship

When recording the character social network relationships, the relationships between characters

are recorded as a triple dimensional matrix that contains two character names and their corresponding relationship attribute and can be optionally exported as a CSV (Comma Separated Values) file of Microsoft Excel or a RDF (Resource Description Framework) file shown as Fig. 15 so that a humanist can further conduct social network analysis via a social network analysis software such as UCINET or Pajek. Therefore, the character social network relationships with RDF format generated by the CSNRMT can easily be published as LD to provide machine-readable information to support Web service in the future.
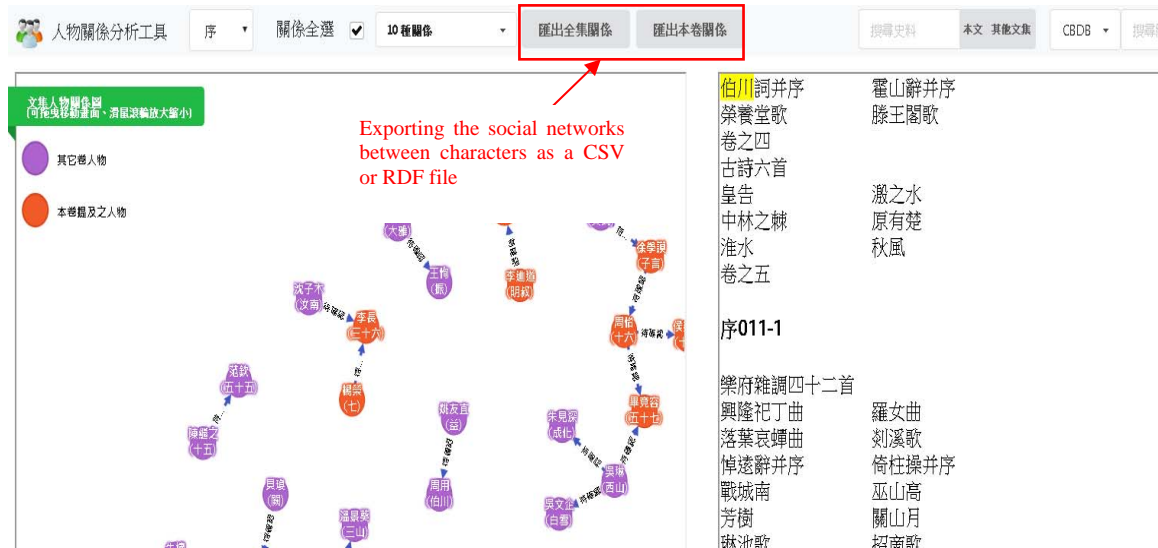

Figure 15. The user interface of exporting character social network relationship

4. Research Methodology

4.1 Research design

To examine the effects of the ATAS and CSNRMT developed on the CABDHRP on assisting digital humanities research, two experiments was designed in this study to examine the effectiveness of the research participants on interpreting Chinese ancient text and characters' social network relationships. The first experiment is that the research participants were invited to interpret a paragraph of ancient text of the Ming Dynasty's writings with the support of ATAS within 20 min, and then wrote down their reading abstracts. Besides, a historical expert who is a university professor of Department of History was invited to select an ancient text with a lot of characters' relationships from Ming Dynasty collection archived in the CABDHRP as the experimental text. The second experiment is that the research participants were invited to read an ancient text with the assistance of CSNRMT for interpreting characters' social network relationships within 40 min, and then wrote down the social relationships of interpreting characters. The semi-structured in-depth interview was preceded after the research subjects completed the reading abstracts and the interpretation of characters' social network relationships. The interview questions include "Are there any difficulties or questions when using the functions of ATAS and CSNRMT in the system", "Is the overall system design clear and understandable", "Is the human-computer interaction design in the system perceived ease of use", and "Are there any opinions or suggestions for the future development of the system".

4.2 Research subjects

The research participants are the students who are able to interpret the ancient texts of Ming Dynasty and explore characters' social network relationships. In consideration of cost, time, and

location, a total of 21 students from Departments of Chinese Literature and History in a national university in Taipei City, Taiwan, who were willing to participate in the experiment, were sampled to conduct two experiments respectively with the support of ATAS and CSNRM for a target task related to digital humanities research. One task was to interpret a paragraph of an ancient text of the Ming Dynasty's writings with the support of ATAS, and then wrote down the reading abstracts. Another one is to read the ancient text with the assistance of CSNRMT for interpreting characters' social network relationships. The 21 students contain 3 undergraduate students, 14 graduate students, and 4 doctoral students.

5. The Use Assessment of ATAS and CSNRM

5.1 The use assessment of ATAS

To understand the research participants' using perception when interpreting the texts with the support of ATAS, the semi-structured interviews with the 21 research participants were preceded. The results are summarized as follows. First, this study found that different opinions appear on the function of the automatic Chinese word segmentation. Most of the research participants agreed with the accuracy of automatic annotation based on Chinese word segmentation in interpreting the contents of the ancient texts and accelerating the reading speed. Other research participants considered that the automatic Chinese word segmentation did not have much assistance, as they could segment the terms by themselves or incorrect Chinese word segmentation might affect the text interpretation. Second, the research participants expressed that the function to link to source websites allows them viewing more useful information. Particularly, the complete annotation information would be more easily browsed on the source websites than on the ATAS. Importantly, most of the research participants strongly expressed the necessity of text segmentation when reading ancient texts without punctuations like the texts in Ming Dynasty's writings. They also regarded that the importance of text segmentation is higher than word segmentation to smoothening the reading. Also, a lot of research participants expressed that there are many variations in Chinese ancient texts and suggested that variations dictionary of Ministry of Education of Taiwan could be included for the search, or the pre-processing of variations could be proceeded before uploading texts to websites. Finally, most of the research participants considered that the ATAS should increase LD sources from other databases. They suggested that increasing online databases or dictionaries which are often used for interpreting Chinese ancient texts, such as Chinese dictionary (http://dict.revised.moe.edu.tw/cbdic/), Scripta Sinica Database (http://hanchi.ihp.sinica.edu.tw/ihp/hanji.htm), and Zdic (http://www.zdic.net/), for expanding the LD sources should be considered.

5.2 The use assessment of CSNRM

To understand the research participants' using perception when interpreting the historical character social network relationships with the support of CSNRMT, the semi-structured interviews with the 21 research participants were preceded. The results are summarized as follows. First, most interviewees considered that the CSNRMT developed in this study presents clearly understandable system interface, clear marks among function blocks, and good mutual reference among system functions. Second, the character social network relationship map could present possible characters and their relationships in the text to help users browse the historical characters' relationships before interpreting the text so as to understand the possible relationships mentioned in the text and allow users not missing characters and the relationship in the text interpretation process. However, many interviewees expressed that some characters in other dynasties would be identified as the characters in Ming Dynasty in the character social network relationship map. For instance, the first founder of Song Dynasty was identified as the first founder of Ming Dynasty, that it would seriously affect the

effectiveness to interpret characters' social network relationships in the text. Also, some interviewees indicated that the characters named as Hsiung sir and Chang sir simultaneously appearing in a text could not be identified their network relationship in the currently developed character social network relationship map because they are the short forms of two character names; however, such characters might play critical roles in the character social network relationship map. Moreover, a lot of interviewees indicated that the system could search characters through the internal search and various external search functions, including Wikipedia and CBDB, to largely assist in interpreting characters' social network relationships in a text. The internal text search function provided by the system allows users clicking on a character in the character social network relationship map to rapidly mark the location of the character appearing in a text. It also allows users finding out useful information related to the character through the external search functions for interpreting characters' social network relationships. Finally, the note function and reading interface developed in this system are presented on the same interface, allowing users reading text and making notes at the same time to reduce the switch times between the reading text interface and note interface. They therefore could pay more attention to interpreting characters' social network relationships in a text.

6. Discussion

The Jieba Chinese parser was applied to the ATAS for automatic word segmentation. However, the Jieba Chinese parser is not particularly developed for processing Chinese ancient texts so that the word segmentation of Chinese ancient texts could not achieve an accuracy as high as modern Chinese texts. Although the ATAS offers the function of adding new terms to improve the accuracy of Chinese word segmentation through semi-automatic adjustment of human-computer interaction, the growth of Chinese word segmentation lexicon size is still too slow to significantly increase the accuracy of word segmentation of Chinese ancient text. Namely, developing a Chinese word segmentation system that can identify Chinese ancient texts well is an urgent issue for the ATAS. Currently, there is no such Chinese word segmentation system to be well developed for Chinese ancient texts. To overcome this problem, developing a Chinese word segmentation lexicon to support a Chinese word segmentation system based on analyzing a large amount of Ming Dynasty's ancient texts collected in the developed CABDHRP is a practicable way. This has been considered as our future work. More importantly, most of the research participants strongly expressed the necessity of text segmentation when reading Chinese ancient texts without punctuations like the texts in Ming Dynasty's writings. They regarded that the text segmentation is more important than word segmentation while reading a Chinese ancient text for digital humanities research. Actually, text segmentation will also influence the accuracy of Chinese word segmentation of an ancient text (Pak & Teh, 2016). Therefore, developing a text segmentation scheme with high accuracy for Chinese ancient texts is a more urgent issue than word segmentation, particularly in using the ATAS to support reading Chinese ancient text without punctuations.

Additionally, the character names' identification in the CSNRMT mainly focuses on Ming Dynasty's characters collected in CBDB. That is, the character names are simply identified based on the lexicon of Ming Dynasty's character names from CBDB. Some research participants indicated that some character names would be wrongly identified, thus appearing interference in reading process. Additionally, some other research participants indicated that some characters before Ming Dynasty are not identified by the system, while such characters might be critical in presenting primary characters' social network relationships. Moreover, the problem of cross-dynasty same names might result in mistakes. For instance, an interviewee mentioned that the first emperor of Song Dynasty is identified as the first emperor of Ming Dynasty due to the same posthumous name. Obviously, the characters' social network relationships identified by the CSNRMT should be more accurate and broad to conform to the text characteristics, types, and humanists' needs. Namely, the

character name recognition technology of Ming Dynasty writings should be further improved based on finding out the rules that can correctly identify character names by using advanced machine learning technologies, such as Hidden Markov Model (Bikel et al., 1997) or Conditional Random Field (Gu et al., 2015). Particularly, human-machine cooperation based on active learning (Settles, 2012) could may help improve the accuracy of the established character name recognition technology of Ming Dynasty writings.

Finally, a variety of digital humanities projects around the world are working to build digital tools for facilitating the research of digital humanists, scholars, educators, but these tools are often "one-offs" from a technology perspective. That is, they focus on a single activity or set of activities and function independently from one another. To provide more complete and powerful functions in supporting digital humanities research, integrating the functions of CABDHRP with the digital tools developed by other digital humanities platforms should be considered. However, integrating existing digital tools into the CABDHRP presents high challenges because they are built using a variety of programming languages and system level dependencies, and in most cases they were never designed with interoperability in mind (Bialock et al., 2018). This study strongly appeals that the issues of function interoperability and data exchange between different digital humanities platforms should be considered in the design process. Generally, using open API to provide system's functions is a practicable way to achieve the goal of function interoperability. Furthermore, there have been a lot of data exchange methods that can be used to achieve the goal of data exchange between different digital humanities platforms, such as via files, an SQL database, open linked data, or web services (Baierer et al., 2016).

7. Conclusions and Future Works

In recent years, with the advancement of information and communication technologies (ICTs), the retrieval and use of Chinese ancient books is arising an unprecedented challenge and opportunity. This study successfully developed the CABDHRP, which can archive any subject's Chinese ancient digital writings and provide several emerging digital tools including the ATAS and CSNRMT, to support DH research. Many interviewees indicated that the user interface and functions of the ATAS and CSNRMT developed in this study are practicable in supporting digital humanities research and easy to use, the functions are clearly marked, and the integration among functions is good. The CABDHRP not only provides innovative and forward-looking applications in DH field, but also promotes the public to emerge the learning interest and awareness of Chinese ancient books in the future. More importantly, this study tried to develop a systematic and standardized digital humanities research platform with novel and practicable digital tools on virtual environment for different humanities disciplines with similar but heterogeneous datasets.

Additional studies are warranted. First, text segmentation technology that divides digital text into meaningful sentences is regarded more important than word segmentation that divides digital text into meaningful words when reading Chinese ancient text without punctuations. Therefore, the future work of enhancing the CABDHRP to support DH research more effectively should focus on developing an automatic text segmentation algorithm to insert appropriate punctuations into Chinese ancient texts for promoting reading comprehension performance of humanists. In this work, the deep learning derived from neural networks, such as the bidirectional LSTM (long short-term memory), has a very high potential to be applied for solving this problem (Wang et al., 2017). Second, annotations typically facilitate text reading. Compared to the automatic text annotation system developed in the CABDHRP, manually collaborative annotations of digital texts can get more correct annotated contents from humanists and allow humanists to share their knowledge with each other. Most importantly, the annotated contents can help readers obtain a deeper and broader understanding while reading a text compared to digital content without annotations. Thus, the future work of this study will integrate the manually collaborative reading annotation system with a

reading annotation and interactive discussion scaffold developed in our previous work into the CABDHRP to improve humanists' reading performance for digital humanities research (Chen & Chen, 2014). Third, character name recognition technology of Ming Dynasty writings should be further improved through finding out the rules that can correctly identify character names by using machine learning technologies. Fourth, opinion leaders in two-step flow of communication model were the brokers between mass media and the public as well as were the key person in information transfer (Li et al., 2013). The future work of this study should develop an opinion leader and maximum influencer identification algorithm for the CSNRMT based on centrality, structure hole, or PageRank to help humanists get useful characters' relationship clues for discovering new insights. Finally, developing a dashboard system for real-time analysis of humanists' behavior processes on the CABDHRP should be considered in the future. If the behavior process records of humanists interpreting text by using the CABDHRP can be analyzed in real-time, then the CABDHRP can provide real-time feedback with more effective information for the text interpretation to humanists.

References

Brooke, J., Hammond, A., & Hirst, G. (2015), "Gutentag: An nlp-driven tool for digital humanities research in the project gutenberg corpus", In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, Denver, Colorado, USA, June. Association for Computational Linguistics, pp. 42-47.

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnapp, J. (2012), *"Digital_humanities",* The MIT Press, Cambridge, Massachusetts London, England.

Bikel D., Schwarta R., Weischedel. R. (1997), "An algorithm that learns what's in a name", *Machine Learning*, Vol. 34, pp. 211-231.

Bialock, T., Childress, D., Ikuura, H., & Nagasaki, K. (2018), "Towards unified descriptive practices for Japanese classical texts: TEI, IIIF, and the UCLA Toganoo Collection of Esoteric Buddhism", *Proceedings of the 8th Conference of Japanese Association for Digital Humanities,* pp.78-79.

Baierer, K., Dröge, E., Eckert, K., Goldfarb, D., Iwanowa, J., Morbidoni, C., & Ritze, D. (2016), "DM2E: A linked data source of digitised manuscripts for the digital humanities. *Semantic Web – Interoperability,Usability, Applicability an IOS Press Journal,* Vol. 2014, No. 1, pp. 1-13. doi: 10.3233/SW-160234

Chan, T. C., Pai, P. L., Shaw, S. L., & Fan, I. C. (2015), "Spatiotemporal evolution of market towns in the Jiangnan area during the Ming-Qing dynasties of China", *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, Vol. 48, No. 2, pp. 90-102, DOI: 10.1080/01615440.2014.995783

Cassella, M. (2017), "New trends in academic library partnerships: Academic libraries and digital humanities", In *Proceedings of the IATUL Conferences*, pp. 1-9.

Chen, C. M. & Chen, F. Y. (2014), Enhancing digital reading performance with a collaborative reading annotation system", *Computers & Education*, Vol. 77, pp. 67-81.

Chen, C. M. & Tsay, M. Y. (2017), "Applications of collaborative annotation system in digital curation, crowdsourcing, and digital humanities", *The Electronic Library*, Vol. 35, No. 6, pp. 1122-1140.

Chen, C. M., Chen, Y. T., Hong, C. M., Liao, C. W., & Huan, C. M. (2012), "Developing a Taiwan

Library History Digital Library with reader knowledge archiving and sharing mechanisms based on the DSpace platform", *The Electronic Library*, Vol. 30, No. 3, pp. 426- 442.

Chen, S. P., Huang, Y. M., Hsiang, J., Tu, H. C., Ho, H. I., & Chen, P. Y. (2013), "Discovering land transaction relations from land deeds of Taiwan", *Literary & Linguistic Computing*, Vol. 28, No.2, pp.257-270.

Drucker, J. (2013), "*Intro to digital humanities: Introduction*", UCLA Center for Digital Humanities. Web available at http://dh101.humanities.ucla.edu/?page_id=13, Retrieved January 28, 2018.

Gu, C., Tian, X. P., Yu, J. D. (2015), "Automatic recognition of Chinese personal name using conditional random fields and knowledge base", *Mathematical Problems in Engineering*, 2015.

Harris, M., Levene, M., & Zhang, D. (2014), "The anatomy of a search and mining system for digital humanities", In *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries,* IEEE Press, Piscataway, NJ, USA, pp. 165-168.

Hashimoto, Y. (2014), "SMART-GS Web: A HTML5-powered, collaborative manuscript transcription platform", Japanese Association for Digital Humanities Annual Conference, Ibaraki, Japan.

Haslhofer, B., Robitza, W., & Lagoze, C. (2013), "Semantic tagging on historical maps", In *ACM Web Science*, Paris, France.

Hsiang, J., Chen, S. P., Ho, H. I., & Tu, H. C. (2012), "Discovering relationships from imperial court documents of Qing dynasty", *International Journal of Humanities and Arts Computing*, Edinburgh University Press, Vol.6, No.1-2, pp. 22-41.

Haslhofer, B., Antoine, I., & Rainer, S. (2018), "Knowledge graphs in the libraries and digital humanities domain", Web available at https://arxiv.org/pdf/1803.03198.pdf

Jackson, C. (2017), 'Using social network analysis to reveal unseen relationships in medieval Scotland", *Digital Scholarship in the Humanities*, Vol. 32, No. 2, pp. 336–343.

Kallaher, A. & Gamble, A. (2017), "GIS and the humanities: Presenting a path to digital scholarship with the Story Map app", *College & Undergraduate Libraries,* Vol. 24, No. 2-4, pp. 559-573.

Kaplan, F. (2015), "A map for big data research in digital humanities", *Frontiers in Digital Humanities,* Vol. 2, No. 1, pp. 1-7.

Kamposiori, C. (2017), "The role of Research Libraries in the creation, archiving, curation, and preservation of tools for the Digital Humanities", RLUK Report. Web available at http://www.rluk.ac.uk/wp-content/uploads/2017/07/Digital-Humanities-report-Jul-17.pdf

Kestemont1, M. & Gussem, J. D. (2017), "Integrated sequence tagging for medieval Latin using deep representation learning", *Journal of Data Mining & Digital Humanities, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages*, pp. 1-17.

Krause, S., Hennig, L., Moro, A., Weissenborn, D., Xu, F., Uszkoreit, H., & Roberto Navigli, R. (2016), "Sar-graphs: A language resource connecting linguistic knowledge with semantic relations from knowledge graphs", *Journal of Web Semantics*, Vol. 37–38, pp. 112-131.

Li, Y., Ma, S., Zhang, Y., & Huang, R., & Kinshuk. (2013), "An improved mix framework for opinion leader identification in online learning communities", *Knowledge-Based Systems*, Vol.

43, pp. 43-51.

Luczak-Roesch, M., Grener, A., & Fenton. E. (2018), "Twenty thousand leagues above the book: An interactive visual analytics approach to literature", In *Proceedings of the 2018 ACM Conference on Supporting Groupwork*, ACM, New York, NY, USA, pp. 94-97.

Martin-Rodilla, P. & Gonzalez-Perez, C. (2016), "Understanding user behavior in textual analysis: A thinking aloud approach for digital humanities research contexts", *The 4th International Conference on Technological Ecosystem for Enhancing Multiculturality*, pp. 269-276.

Neuroth, H., Lohmeier, F., Smith, K. M. (2011), "TextGrid—virtual research environment for the humanities", *International Journal Digital Curation*, Vo. 6, No. 2, pp. 222-231.

Pastorelli, D. (2017), "A classification of manuscripts based on a new quantitative method", The Old Latin witnesses of John's Gospel as text case. *Journal of Data Mining & Digital Humanities, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages*, pp. 1-48.

Picca, D., Egloff, M. (2017), "DHTK: The digital humanities ToolKit", CEUR Workshop Proceedings, *2014*, pp. 81-86.

Pak, I. and Teh, P. L. (2016), "Text segmentation for analysing different languages", *The First EAI International Conference on Computer Science and Engineering*, 11-12 November 2016, Penang, Malaysia.

Rushmeier, H. Pintus, R. Yang, Y. Wong, C. and Li, D. (2015), "Examples of challenges and opportunities in visual analysis in the digital humanities", web available at: http://graphics.cs.yale. edu/site/sites/files/dh-ei-SUBMIT.pdf.

Sato, T, Goto, M., Kimura, F., & Maeda, A. (2016), "Developing a collaborative annotation system for historical documents by multiple humanities researchers", *International Journal of Computer Theory and Engineering,* Vol. 8, No. 1, pp. 88-93.

Steiner, C.M., Agosti, M., Sweetnam, M. S., Hillemann, E. C., Orio, N., Ponchia, C., Hampson, C., Munnelly, G., Nussbaumer, A., Albert, D., & Conlan, O. (2014), "Evaluating a digital humanities research environment: the CULTURA approach", *International Journal on Digital Libraries*, Vol. 15, No. 1, pp. 53-70.

Settles, B. (2012), "Active learning", *Synthesis Lectures on Artificial Intelligence and Machine Learning*, Vol. 6, No. 1, pp. 1-114.

Tang, M. C., Cheng, Y. J. & Chen, K. H. (2017), "A longitudinal study of intellectual cohesion in digital humanities using bibliometric analyses", *Scientometrics*, Vol. 113, No. 2, pp. 985-1008.

Therón, R. & Wandl-Vogt, E. (2016), "New trends in digital humanities", In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM '16)*, Francisco José García-Peñalvo (Ed.). ACM, New York, NY, USA, pp. 945-947.

Tin Can API. (2015), "*What is the Tin Can API?",* web available at https://tincanapi.com/overview/

Uboldi, G., Caviglia, G., Coleman, N., Heymann, S., Mantegari, G., & Ciuccarelli, P. (2013), "Knot- an interface for the study of social networks in the humanities", In *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI (CHItaly '13)*, ACM, New York, NY, USA, Article 15, 9 pages.

Upadhyay, S. & Upadhyay, N. (2017), "Future directions and a roadmap in digital computational humanities for a data driven organization", *Procedia Computer Science*, Vol. 122, pp. 1055-1060.

Wang, X., & Inaba, M. (2009), "Analyzing structures and evolution of digital humanities based on correspondence analysis and co-word analysis",アート・リサーチ, Vol. 9, pp. 123-134.

Wang, X., Wang, M., & Zhang, Q. (2017), "Realization of Chinese word segmentation based on deep learning method", In *Proceedings of AIP Conference*, 1864, 020150-1–020150-6.

Widlöcher, A., Bechet, N., Lecarpentier, J. M., Mathet, Y., & Roger, J. (2015), "Combining advanced information retrieval and text-Mining for digital humanities", *The ACM Symposium on Document Engineering*, Lausanne, Switzerland, pp. 157-166.

Yohan, P. M., Sasidhar, B., Basha, Sk. A. H., & Govardhan, A. (2014), "Automatic named entity identification and classification using heuristic based approach for Telugu", *International Journal of Computer Science Issues*, Vol. 11, No. 1, pp. 173-180.