Development and evaluation of an automatic text annotation system for supporting digital humanities research

Chih-Ming Chen, Yung-Ting Chen and Chen-Yu Liu
National Chengchi University, Taipei, Taiwan

Received 31 October 2017
Revised 28 May 2018
10 August 2018
7 September 2018
Accepted 6 October 2018

Abstract

Purpose – An automatic text annotation system (ATAS) that can collect resources from different databases through Linked Data (LD) for automatically annotating ancient texts was developed in this study to support digital humanities research. It allows the humanists referring to resources from diverse databases when interpreting ancient texts as well as provides a friendly text annotation reader for humanists interpreting ancient text through reading. The paper aims to discuss whether the ATAS is helpful to support digital humanities research or not.

Design/methodology/approach – Based on the quasi-experimental design, the ATAS developed in this study and MARKUS semi-ATAS were compared whether the significant differences in the reading effectiveness and technology acceptance for supporting humanists interpreting ancient text of the Ming dynasty's collections existed or not. Additionally, lag sequential analysis was also used to analyze users' operation behaviors on the ATAS. A semi-structured in-depth interview was also applied to understand users' opinions and perception of using the ATAS to interpret ancient texts through reading.

Findings – The experimental results reveal that the ATAS has higher reading effectiveness than MARKUS semi-ATAS, but not reaching the statistically significant difference. The technology acceptance of the ATAS is significantly higher than that of MARKUS semi-ATAS. Particularly, the function comparison of the two systems shows that the ATAS presents more perceived ease of use on the functions of term search, connection to source websites and adding annotation than MARKUS semi-ATAS. Furthermore, the reading interface of ATAS is simple and understandable and is more suitable for reading than MARKUS semi-ATAS. Among all the considered LD sources, Moedict, which is an online Chinese dictionary, was confirmed as the most helpful one.

Research limitations/implications – This study adopted Jieba Chinese parser to perform the word segmentation process based on a parser lexicon for the Chinese ancient texts of the Ming dynasty's collections. The accuracy of word segmentation to a lexicon-based Chinese parser is limited due to ignoring the grammar and semantics of ancient texts. Moreover, the original parser lexicon used in Jieba Chinese parser only contains the modern words. This will reduce the accuracy of word segmentation for Chinese ancient texts. The two limitations that affect Jieba Chinese parser to correctly perform the word segmentation process for Chinese ancient texts will significantly affect the effectiveness of using ATAS to support digital humanities research. This study thus proposed a practicable scheme by adding new terms into the parser lexicon based on humanists' self-judgment to improve the accuracy of word segmentation of Jieba Chinese parser.

Practical implications – Although some digital humanities platforms have been successfully developed to support digital humanities research for humanists, most of them have still not provided a friendly digital reading environment to support humanists on interpreting texts. For this reason, this study developed an ATAS that can automatically retrieve LD sources from different databases on the Internet to supply rich annotation information on reading texts to help humanists interpret texts. This study brings digital humanities research to a new ground.

Originality/value – This study proposed a novel ATAS that can automatically annotate useful information on an ancient text to increase the readability of the ancient text based on LD sources from different databases, thus helping humanists obtain a deeper and broader understanding in the ancient text. Currently, there is no this kind of tool developed for humanists to support digital humanities research.

Keywords Digital humanities, User behaviour, Linked Data, Automatic segmentation of Chinese word, Automatic text annotation system, Reading interface design

Paper type Research paper

Library Hi Tech
Vol. 37 No. 3, 2019
pp. 436-455

0737-8831
DOI 10.1108/LHT-10-2017-0219

The authors would like to thank the Research Center for Chinese Cultural Subjectivity of National Chengchi University for financially supporting this research under Contract No. 107H21.

1. Introduction

Since the initiation of Digital Archives Program in 2002, a lot of academic institutes in Taiwan have digitalized the important archives. Although a large amount of data have been accumulated in the past decade, most of such digital archive databases are independent and cannot be integrated for the utilization. Besides, most humanists stay the imagination of digital humanities at the stages of digital archives or the digitalization of historical data, rather than thoroughly utilizing such resources for deeper research. Rosenzweig (2003) indicated that a researcher did not encounter the lack of data but how to deal with excessive data; therefore, how to make such data appear meanings was the problem for digital humanities. Moreover, the text reading environment to support digital humanities research is currently short. As the example of Taiwan History Digital Library (THDL) (Hsiang *et al.*, 2009) (http://thdl.ntu.edu.tw/index.html), the database covers more than a hundred thousand full-text data of Tan-Hsin Archives, Ming and Qing Archives of Taiwan Administration, and Ancient Contracts, but the digital library stresses on the development of data analysis tools and is lacking a friendly data interpretation reader for humanists. Most humanists therefore simply utilize the database for data search so that the benefit to support digital humanities research is reduced. Another platform, CBETA Research Platform (CBETA-RP) (http://cbeta-rp.dila.edu.tw/), provides an online reader for Chinese Buddhist texts, with complete contents. It currently also provides researchers with reference of names; however, there is merely mutual reference of internal data, and the integration with cross-platform resources is insufficient (Tu *et al.*, 2012).

To offer a digital humanities reading environment which could integrate cross-platform resources, provide a friendly reader and digital tools for effectively assisting humanists in digital humanities research, Scheinfeldt (2010) pointed out the similarity between a digital humanities scholar and a scientist that both of them extremely depended on tools. A new digital tool could solve the past humanities research problems. Monte and Serafin (2017) indicated that the first and most salient theme that emerged in digital humanities research is the requirement of digital reading and research tools. To effectively support digital humanities research, 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 embed with any HTML web pages to enable users to collaboratively add and manage annotations on HTML web pages and provide a shared mechanism for discussing about shared annotations among multiple users. By applying the CAS in mashup on static HTML web pages, their study discussed the potential applications of CAS in digital humanities. However, the CAS is a kind of manual annotation system. The quality of annotations from users may not be qualified enough to support digital humanities research. Moreover, MARKUS semi-automatic text annotation system is an online text reading and research tool developed by Ho and Hilde (2014) for supporting digital humanities research. A user could upload texts and select the required annotation types in MARKUS, which would then annotate the terms in the text as well as provide the user with data search on Wikipedia, China Biographical Database (CBDB), Temporal Gazetteer (TGAZ) and ZDict, to help the user interpret the text content online. However, the annotation function of MARKUS is limited only for annotating the predefined terms, including personal names, place names, temporal references and bureaucratic offices in a text because of the absence of the function of automatic segmentation of word, thus likely reducing the effectiveness of supporting humanists to interpret the text. As a result, an automatic text annotation system (ATAS) for supporting digital humanities research was developed in this study to collect resources from different databases, through LD, and automatically annotate texts for the users real-time referring to resources from different databases when interpreting texts. Besides, a friendly text annotation reader is provided for humanists interpreting the data through reading. This study aims to confirm whether the

proposed ATAS provides benefits in promoting the reading comprehension of humanists and has high technology acceptance. Additionally, the operation behaviors of humanists while using the proposed ATAS for interpreting an ancient text were also explored based on lag sequential analysis (Bakeman, 1986).

2. Literature review

2.1 Current development of digital humanities research platform

This study surveyed several current digital humanities research platforms in Taiwan and around the world, including CULTURA (www.cultura-strep.eu/outcomes#2), Scripta Sinica Database (http://hanchi.ihp.sinica.edu.tw/ihp/hanji.htm), THDL (http://thdl.ntu.edu.tw/index.html), CBETA-RP (http://cbeta-rp.dila.edu.tw/) and MARKUS semi-ATAS (http://dh.chinese-empires.eu/markus/beta/). The CULTURA is a corpus agnostic research environment integrating innovative service that guides, assists and empowers a broad spectrum of users in their interaction with cultural artefacts (Steiner *et al.*, 2014). 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. 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). 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 statistic of term, and relevant bibliography are also provided in the CBETA-RP, whereas MARKUS is a famous semi-ATAS and supports online text reading.

Table I shows the function comparison of the five digital humanities platforms. It was discovered that Scripta Sinica Database presents the richest full-text data collection, CBETA-RP merely includes Buddhist texts, and MARKUS semi-ATAS does not show the full-text data collection function. The CULTURA system consists of multiple distinct services including personalized search tools, faceted search tools, annotators, social network

Function comparison of digital humanities platforms	CULTURA	Scripta Sinica Database	Taiwan History Digital Library	CBETA Research Platform	MARKUS semi-automatic text annotation system
Full-text data collection	O	O	O	Merely data of Buddhist texts	X
Automatic segmentation of Chinese word	X	X	X	X	X
Automatic annotation	O	X	O	X	O
Term search	O	O	O	O	O
External reference	O	Qing Officials Query System	Variants Database	Database of Buddhist Tripitaka	CBDB and Zdict
Word frequency statistics	O	X	O	O	X
Bookmark notes	X	X	O	X	Adding annotations
Social network visualization tools	O	X	X	X	X
Recommender	O	X	X	X	X

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

Table I.
Function comparison of digital humanities research platforms

visualization tools and recommenders (Steiner *et al.*, 2014). Besides, all platforms are lacking automatic segmentation of Chinese words; the automatic annotation function exists in the CULTURA, THDL and MARKUS semi-ATAS, and each system presents the search function. In terms of external reference, Scripta Sinica Database could connect to Qing Officials Query System, THDL presents variant databases, CBETA-RP merely includes Buddhist text databases and MARKUS semi-ATAS shows the most external reference resources, containing Wikipedia, CBDB, TGAZ and ZDict. Merely CULTURA, THDL and CBETA-RP show the function of word frequency statistics; merely THDL presents the function of bookmark notes; and MARKUS semi-ATAS has the function to add annotations.

From above statements, although some digital humanities research platforms could be used to support digital humanities research for humanists, the system functions of each platform are rather insufficient, and a friendly digital reader has still not been offered. For this reason, it is expected to develop an ATAS for supporting digital humanities research. Resources from different databases are integrated as the annotation information to help humanists interpret texts based on LD. Besides, an embedded digital tool was developed for integrating other text reading platforms which could support digital humanities research to provide a digital humanities research environment for the easy use of humanists. Accordingly, it is expected to have humanists gradually perceive the benefits of information technology to research and allow more people engaging in digital humanities research.

2.2 Applications of Linked Data

Linked Data (LD) is a kind of structured data published on the web following a set of principles designed to promote the interlinking between the various data sets on the web (Dutta, 2017). The development of LD could be traced back to Tim Berners-Lee, the founder of World Wide Web (WWW) (Bizer *et al.*, 2009). Following the flourishing of Internet and WWW, more and more data are uploaded to the Internet and the Internet has become an important channel for acquiring information. However, it is another problem to effectively search, integrate, and utilize such rapidly growing resources. Berners-Lee and Fischetti (1999) therefore proposed the idea of semantic Web (Web3.0). The past Internet was written with HTML web pages and presented with browsers, and the data were made with the format that is convenient for people's reading. However, a computer could not understand the meanings of data. The idea of semantic Web was to transfer data into an interpretable form for computers so as to share, integrate, and utilize resources on the Internet. To realize semantic Web, Berners-Lee *et al.* (2006) further proposed the idea of LD, expecting to structure the data on WWW and allow such data connecting to each other on the Internet and organizing information with same ideas.

Auer *et al.* (2007), the researchers of Universität Leipzig, Freie Universität Berlin and University of Pennsylvania indicated that a large amount of data on Wikipedia was frequently browsed, but the data, due to the restriction on format, could hardly be linked with other databases for different applications. A program to extract resources from Wikipedia, transfer to the structured format of LD and save as a data set for open connection and utilization, called DBpedia, was therefore developed. The British Broadcasting Corporation (BBC), on the other hand, linked internal data among different databases through LD. BBC operate several radio stations and TV channels, in which the content management systems are independent so that the resource sharing is not convenient. They therefore attempted to use the data set of DBpedia and MusicBrainz as the controlled terms and connect the contents with the same topic in databases of different systems so that all channels and radio station websites under BBC could mutually share and link the resources (Kobilarov *et al.*, 2009). Libraries have several applications of LD. For example, Virtual International Authority File is the authority file data with RDF format cooperatively established by Library of Congress, Die Deutsche Nationalbibliothek,

Bibliothèque nationale de France and OCLC to provide international authority files which could be mutually shared (Bennett *et al.*, 2006). WorldCat of OCLC developed LD of bibliographic records, allowing the bibliographic data from various libraries being mutually connected. It provides a good basis for innovative reader service and broad communication of metadata (Cole *et al.*, 2013).

In sum, the idea of LD has developed many value-added applications for the Internet data from different sources being mutually shared and connected. It facilitates more effective utilization of resources from the Internet and solves the situation of independent digital archive databases not being able to precede value-added applications. As a consequence, the idea of LD was applied to the development of ATAS for supporting digital humanities in this study. With the characteristic of LD being able to integrate data, resources from different databases, it is suitable to become the annotation information source for automatic annotation to assist humanists in interpreting texts.

3. The developed ATAS

3.1 System architecture of ATAS

Figure 1 shows the system architecture of the proposed ATAS in this study. According to the procedures marked in the system architecture diagram, the operation of the system is explained as follows:

(1) Since there is not a boundary mark between Chinese words, it is necessary to pre-process texts before word segmentation. In the pre-processing of word segmentation, the developed system would automatically separate texts with punctuation into sentences before the formal word segmentation, and then input them to the parser for word segmentation.

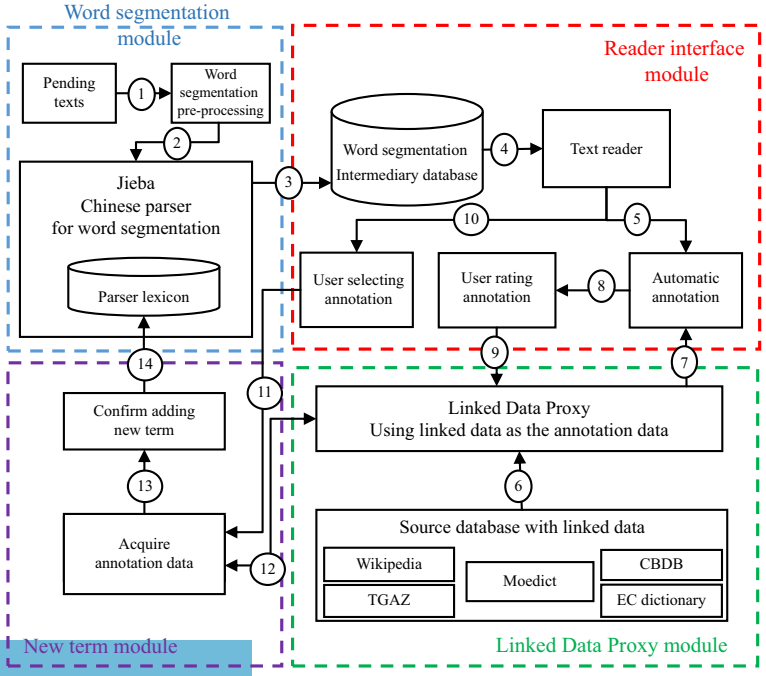

Figure 1.
System architecture of the proposed ATAS

(2) Jieba Chinese parser is utilized in this study, and a self-defined lexicon, in which the term could be expanded, is used. Based on all possible wordings in the sentences in the tree structure, a dynamic planning is used for finding out the path with the maximal probability. Such a path is the word segmentation result based on word frequency. At this stage, the term in the parser lexicon would be searched for the most correct word segmentation till the entire text completes the segmentation of word.

(3) The terms acquired from word segmentation of texts at the previous stage are stored in the word segmentation term intermediary database for text presentation and automatic annotation.

(4) Texts with word segmentation are output to the text reader for presenting the full text.

(5) After loading texts with word segmentation into the reader, the system starts loading automatic annotation and automatically annotating the segmented terms.

(6) LD Proxy is utilized for the automatic annotation of terms. Taking LD as the annotation source, the acquired LD from different databases would be loaded as reference data for automatic annotation.

(7) The reference data for annotated terms would be acquired from the databases including Wikipedia, CBDB, TGAZ, Moedict and EC dictionary. The databases of Wikipedia and TGAZ contain LD for direct data acquisition, while the data from other databases need to be transformed with the LD Proxy module.

(8) After completing the automatic annotation process, a user could move the cursor on the reader to the term with automatic annotation and click for the annotation data from various linked databases, which are orderly arranged according to linked databases. In this case, a user could judge the correctness and helpfulness of annotation and rate annotation data from such databases.

(9) After the user rates the annotations from various databases, the system would calculate such rating scores, reload automatic annotations, and order the data from different databases, according to the annotation scores, for the term; the LD with high annotation scores are displayed in priority.

(10) Since Jieba Chinese parser is based on a parser lexicon, some terms in the text reader might appear wrong word segmentation or unknown terms. A user could select unknown terms missed in the lexicon as new terms for automatic annotation.

(11) After selecting an unknown term as a new term, the system would acquire the annotation data for the term and provide the user with reference to judge whether the term is a new one or not.

(12) The system would transmit the unknown term, which requires annotation, to the LD Proxy module to search LD for supplying the annotation information.

(13) The user selects the unknown term, which requires annotation, and refers the annotation information. When the selected unknown term is confirmed as a new term, the user can click the “Add new term” button to add the term into the parser lexicon for increasing the accuracy of Chinese word segmentation.

(14) After adding a new term at the previous stage, the system would automatically add the term to the parser lexicon. When the term appears in other paragraphs or texts, the correct word segmentation would be generated for automatic annotation so as to provide readers with an automatic annotation for supporting digital humanities research.

3.2 System implementation of ATAS

The ATAS developed in this study used the Node.js framework to develop the system's front-end and back-end services. Using the Node.js framework to develop web-based systems not only can allow the front-end and back-end programs to be written in a more consistent and rapid manner, but also can utilize many third-party modules provided by its vast development community to greatly reduce system development time. In addition, using the Node.js to develop web-based systems also allows future systems to have better compatibility with different platforms. The LD sources of the ATAS contain five databases, including Wikipedia, CBDB, TGAZ, Moedict and English-Chinese dictionary. Among the five databases, CBDB and TGAZ both provide Application Programming Interface (API) so that the ATAS can easily communicate with these two source databases and obtain the necessary open resources. However, databases such as Wikipedia, Moedict and English-Chinese Dictionary do not provide API. Therefore, Web API which is an HTTP service framework was used to obtain LD with JSON or XML format through Uniform Resource Identifier by using GET or POST commands. Moreover, PostgreSQL was used as a database to store the results of word segmentation from Jieba Chinese parser and LD from the different source databases in the ATAS. The third-party module "pg" in the Node.js was used to perform the data reading or writing operations between the ATAS and PostgreSQL. Currently, the ATAS is still a digital humanities research tool and is available at: (http://exp-linked-data-proxy-2017.dlll.nccu.edu.tw:3253/directory_ming). In the future, the ATAS will be further applied in a Chinese ancient digital humanities research platform to support digital humanities research and scholarship.

3.3 System interface and functions

The user interface and functions of the proposed ATAS for supporting digital humanities are explained as follows:

(1) Highlight automatic annotated word: when a digital humanities researcher enters the system for text reading, the system would perform word segmentation and display the segmented terms by using blue highlight. When the researcher moves the cursor to the segmented term with automatic annotation, the system would change the term's color as red highlight to remind that the term has available annotation information for the researcher (Figure 2).

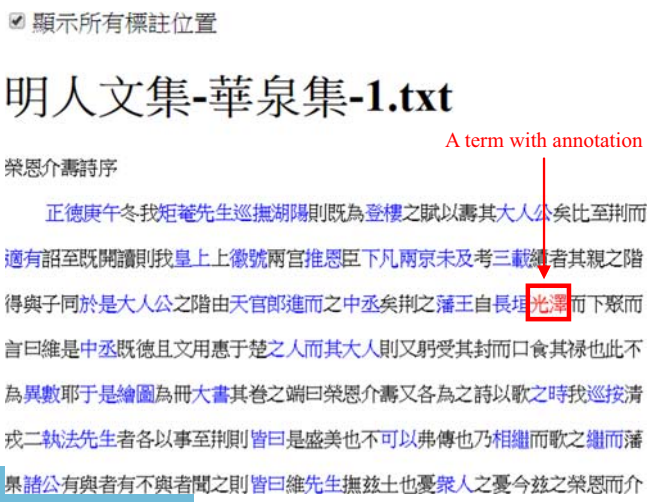

Figure 2.
Highlighting a term with annotation

(2) Select source database for viewing annotation content: the system would present the annotation content after the researcher clicks on the term with automatic annotation. On 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 database for the term (Figure 3).

(3) Link to source website: the automatic annotations from the LD of different databases merely extract partial contents for the researcher's reading reference. The researcher could also click on the "Reference button" below the annotation content, which is linked to the original data source website, for getting more information (Figure 4).

(4) Rate data source with or without helpfulness in understanding: when a researcher clicks to view the automatic annotation content, the annotations from different databases are displayed with default order in the system. The user could rate such annotations according to the helpfulness or not. The user would enhance the annotation score for a useful database; on the other hand, the helpless one would reduce the annotation score. The system would calculate the annotation scores of different databases. When the researcher re-clicks the annotation, the annotations from different databases would be reordered according to the annotation scores; ones with higher annotation scores are displayed in priority (Figure 5).

Figure 3.
Selecting source database for viewing annotation content

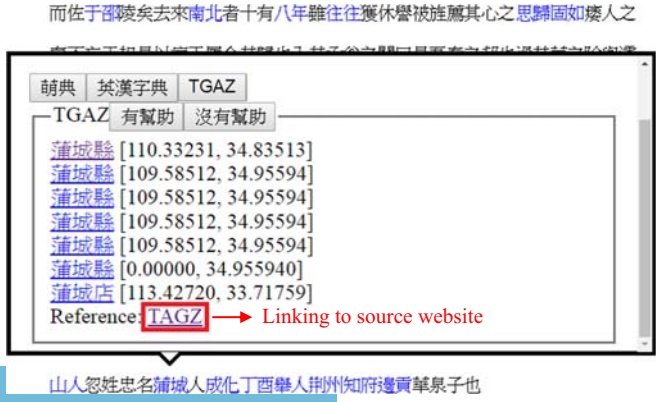

Figure 4.
Linking to source website for viewing more information

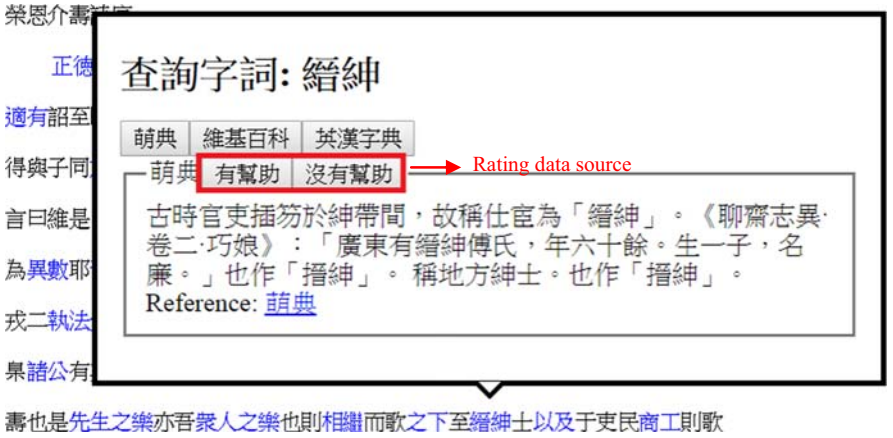

Figure 5.
Rating data source with or without helpfulness in understanding

(5) Add an unknown term as a new term: the performance of automatic annotation would be affected by the correctness of word segmentation. Generally, some terms might not show automatic annotations because they are unknown terms to the used Chinese word segmentation system. In this case, the researcher could select an unknown term in the ATAS system, and then the researcher could consider adding the term as a new term when there are useful LD from the considered databases for the term. After clicking on the "Add new term button," the system would pop out a window for confirmation, and then the term is added to the parser lexicon of the Chinese word segmentation system Jieba. The system would precede correct word segmentation and automatic annotation of the terms in the next loading to the researcher (Figure 6).

Compared to the five digital humanities platforms shown in Table I, the proposed ATAS is the unique one with automatic text annotation functions supported by a Chinese parser with manually adding new terms. Additionally, the proposed ATAS can give lower ranking for the LD resources that are not highly related to the annotated terms based on collective intelligence from humanists who rate the LD resources with or without helpfulness by their professional judgment. Therefore, the proposed ATAS can provide richer and more correct annotations to help humanists understand Chinese ancient texts than other five digital humanities research platforms.

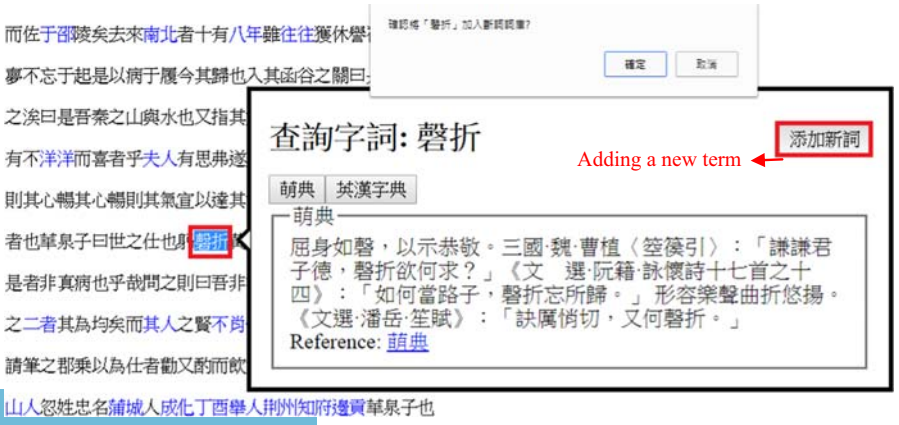

Figure 6.
Adding an unknown term as a new term

4. Research methodology

4.1 Research participants

The research participants are the students who are able to interpret the ancient texts of Ming dynasty and write reading abstracts. In consideration of cost, time and location, a total of 31 undergraduates or graduates of Departments of Chinese Literature and History in a national university in Taipei City, Taiwan, who were willing to participate in the experiment, were sampled. The 31 students contain 11 undergraduates and 8 graduates of Department of Chinese Literature, 1 graduate of Graduate Institute of Taiwan Literature, 8 undergraduates and 1 graduate of Department of History, 1 graduate of Graduate Institute of Taiwan History and 1 undergraduate of Department of Public Administration.

4.2 Experiment design and procedure

In the experiment, the participants interpreted two paragraphs of the Ming dynasty's collections with the support of ATAS and MARKUS semi-ATAS, and then write the reading abstracts. They were also requested to fill in the technology acceptance model questionnaire and precede a semi-structured in-depth interview. The significant differences in technology acceptance and reading effectiveness on the text interpretation for supporting digital humanities between the ATAS and MARKUS semi-ATAS were compared in this study. Before the experiment, the two paragraphs of Ming dynasty's collections were confirmed that they have the close difficulty in interpretation by the experts of Ming dynasty's collections. The experimental processes in this study are planned as Figure 7. The experimental processes are divided into two stages and the total experiment time is 120 min. To ensure the smooth experiment, the experimental procedures were explained before performing the experiment to let the research participants understand the experimental objectives and the experimental processes. The operation of the two systems was further taught at two stages. After understanding the system's operation, the research participants have to respectively interpret the texts with the support of ATAS and MARKUS semi-ATAS within 40 min. The research participants who used ATAS could browse any segmented terms with annotation in the experimental processes. Aiming at unknown terms in the ancient texts, the research participants who used ATAS, according to the knowledge background and the annotations provided by the ATAS, could make judgment to add as new terms into the parser lexicon of Jieba Chinese parser. The experiment was not interfered so that the research participants could freely interpret the ancient texts. The text abstracts were written at the same time; whether the research participants could effectively interpret the ancient texts was assessed through the text abstracts.

To prevent the experiment results from the influence of the sequence of using the systems and reading the texts, the 31 research participants were divided into four groups and used the two systems and two texts in a crisscross pattern at stage 1 and stage 2. Table II shows the adopted experiment system and sequence of text use.

4.3 Research tools

4.3.1 MARKUS semi-ATAS. MARKUS semi-ATAS aims to automatically annotate the terms of personal names, place names, temporal references, and bureaucratic offices in a text and provide the user with data search on Wikipedia, CBDB, TGAZ and ZDict to help the user interpret the text content online. The functions of MARKUS and the proposed ATAS in this study for supporting humanists to interpret the texts are compared and shown in Table III. It was discovered that MARKUS lacks the functions of full-text data collection and automatic segmentation of word, while the ATAS does not contain the functions of searching the location of a term in the text and annotation classification. Obviously, due to without the function of automatic segmentation of word, the annotation function of MARKUS is limited

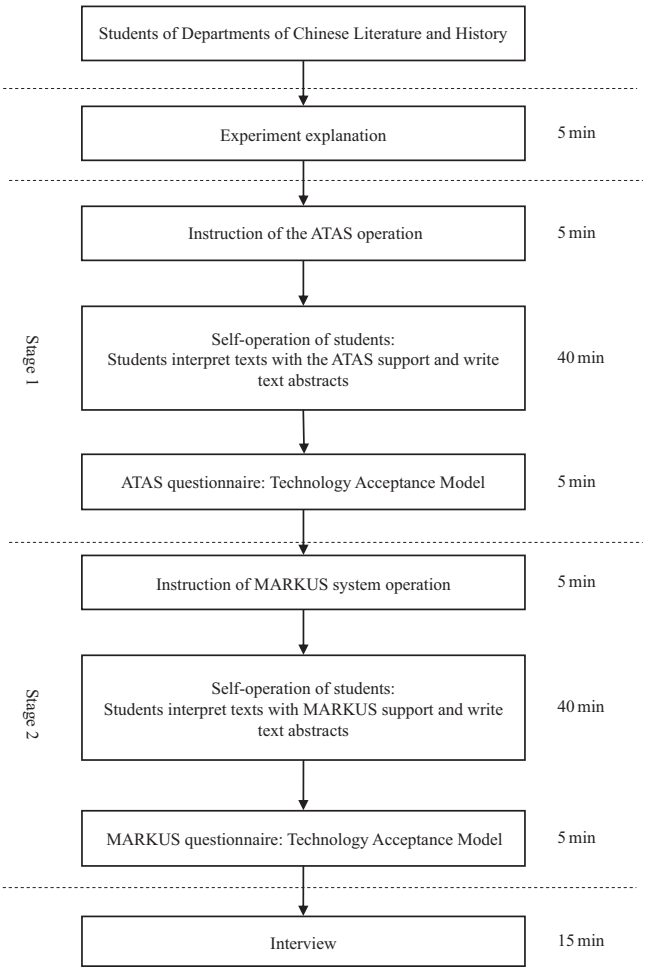

Figure 7.
The experimental procedure of the study

Table II.
The adopted experiment system and sequence of text use

	Stage 1	Stage 2	Number of users
Group 1	ATAS (text 1)	MARKUS (text 2)	8
Group 2	ATAS (text 2)	MARKUS (text 1)	8
Group 3	MARKUS (text 1)	ATAS (text 2)	7
Group 4	MARKUS (text 2)	ATAS (text 1)	8

only for the predefined terms, including personal names, place names, temporal references and bureaucratic offices in a text, thus reducing the effectiveness of supporting humanists to interpret the texts. In regard to the external references supported by the both systems, merely Moedict and Zdict show differences.

4.3.2 Technology acceptance model questionnaire. To understand the research participants' opinions about the ATAS and MARKUS for supporting digital humanities

research, the participants were invited to fill in the technology acceptance model questionnaire after ending the experiment. The analysis of the technology acceptance model could help understand the research participants' subjective perception of using the ATAS and MARKUS semi-automatic annotation system to assist the text comprehension as well as the perceived difficulty of the research participants in operating such two systems.

Referring to the technology acceptance model compiled by Hwang *et al.* (2013) and revising some sentences to conform to the research requirement, a Likert six-point scale was used for the marking. The model contains two dimensions of perceived usefulness of system with 6 questions and perceived ease of system use with 7 questions, totalling 13 questions. Regarding the reliability, the Cronbach's αs of perceived usefulness of system and perceived ease of system use are 0.95 and 0.94, respectively; both present good reliability.

5. Experimental results

5.1 Comparison of the number of annotations automatically generated by the ATAS and MARKUS

Table IV shows the comparison of the number of annotations automatically generated by the ATAS and MARKUS for the two paragraphs of Chinese ancient text of the Ming dynasty's collections used in this study. The results show that the number of annotations automatically generated by the ATAS is much higher than the MARKUS. In other words, the ATAS with the support of Jieba Chinese parser and manually adding new terms can provide richer annotations based on LD to help humanists interpret Chinese ancient texts than does the MARKUS.

5.2 Analysis of writing effectiveness of text abstract

In the experiment, the 31 research participants read four paragraphs of Ming dynasty's collections with the support of ATAS and MARKUS semi-automatic annotation system and write the reading abstracts. Experts were invited to mark the abstracts. The full mark for the abstract of each paragraph is 5 and total 20 are made for the four paragraphs. Independent-samples *t*-test was further used for comparing the difference in the research

System function	ATAS	MARKUS
Full-text data collection	O	X
Automatic segmentation of word	O	X
Automatic annotation	O	O
Term search	O	O
External reference	Moedict, CBDB, TGAZ, Wiki	Zdict, CBDB, TGAZ, Wiki
Searching the location of a term in text	X	O
Bookmark notes	X	X
Adding vocabulary	O	Adding annotation
Link to source website	O	O
Annotation classification	X	O

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

Table III. Function comparison between the ATAS and MARKUS semi-automatic text annotation system

Comparison item	ATAS	MARKUS
Number of annotations in text 1	315	83
Number of annotations in text 2	336	66
Total number of annotations in texts 1 and 2	651	149

Table IV. Comparison of the number of annotations automatically generated by the ATAS and MARKUS

participants interpreting the texts between two systems. The results are shown in Table V, where the average mark for the abstracts written by the research participants interpreting the texts with the ATAS support is 7.94, which is higher than the average mark 6.87 with MARKUS support. However, the effectiveness of such two systems for assisting research participants in interpreting the texts does not reach the statistically significant difference ($t = 1.044$, $p = 0.301 > 0.05$).

5.3 Analysis of difference in technology acceptance degree

After completing the experiments with two systems, the research participants were requested to fill in the technology acceptance model questionnaire. The total score for technology acceptance degree is the sum of the scores of perceived usefulness and perceived ease of use. The sum of technology acceptance degree of two systems was proceeded independent-samples *t*-test. First, the average scores of the sum of technology acceptance degree of two systems were analyzed the difference. The independent-samples *t*-test results of the average score are shown in Table VI, which reveals that the technology acceptance degree of the ATAS is significantly better than that of MARKUS semi-ATAS ($t = 1.068$, $p = 0.037 < 0.05$). Such a result shows that the research participants appear more positive acceptance on the ATAS developed in this study than the MARKUS semi-ATAS.

To ensure the factors in the difference in humanists' technology acceptance degree during using the two systems, perceived usefulness and perceived ease of use of technology acceptance were examined by using independent-samples *t*-test, respectively. The results are shown in Table VII, which reveals that the means of the participants' perceived usefulness of the ATAS and MARKUS semi-ATAS are 21.23 and 21.55, respectively. Both systems do not achieve the statistically significant difference in the perceived usefulness ($t = -0.204$, $p = 0.839 > 0.05$).

The results are shown in Table VIII, which shows that the perceived ease of use of the ATAS is remarkably superior to that of MARKUS semi-ATAS ($t = 2.038$, $p = 0.046 < 0.05$). Such a result shows that the participants considered that interpreting the texts with the ATAS support is easier than with the MARKUSS semi-ATAS.

Table V.
Independent-samples *t*-test of text interpretation effectiveness for both systems

System	Number of users	Mean	SD	*t*	Sig. (two-tail)
ATAS	31	7.94	4.312	1.044	0.301
MARKUS	31	6.87	3.695		

Table VI.
Independent-samples *t*-test of the technology acceptance for both systems

System	Number of users	Mean	SD	*t*	Sig. (two-tail)
ATAS	31	55.52	7.206	1.068	0.037
MARKUS	31	52.94	11.361		

Table VII.
Independent-samples *t*-test of perceived usefulness of technology acceptance of two systems

System	Number of users	Mean	SD	*t*	Sig. (two-tail)
ATAS	31	21.23	5.608	−0.204	0.839
MARKUS	31	21.55	6.806		

5.4 Analysis of difference for the linked databases used in the ATAS

The linked databases of the ATAS developed in this study contain Wikipedia, Moedict, CBDB, TGAZ and EC dictionary. To understand the assistance of the annotation information acquired from the considered linked databases in supporting the research participants to interpret the texts, the collected five-point scale data are first preceded descriptive statistics. The results are shown in Table IX.

From Table IX, the research participants considered that Moedict in the ATAS shows the best assistance in interpreting the ancient texts, with the average mark 4.06, while EC dictionary appears the lowest assistance, with the average mark 2.23. Moreover, one-way analysis of variance was further preceded to examine whether the significant differences between the five linked databases in assisting ancient text comprehension existed. The results are shown as Table X. The results show that the research participants considered significant differences in the assistance of the linked databases in the ATAS ($F = 18.789$, $p = 0.000 < 0.05$). Scheffe method was then utilized for multiple comparisons. The results reveal the higher assistance of Moedict, Wikipedia, CBDB and TGAZ than EC dictionary, while the rest does not show significant differences.

The above analysis merely shows the difference in the assistance among linked databases. This study intends to further understand the consistent viewpoints of the research participants about the assistance of linked databases. Pearson's χ^2 test was therefore utilized for analyzing the assistance distribution of linked databases and the difference in expected number of times. The results are shown in Table XI, where Moedict presents remarkably higher expected number of times on helpful ($2.1 > 1.96$) and extremely helpful ($2.6 > 1.96$), revealing that most research participants are inclined to the helpfulness of Moedict in the text interpretation. TGAZ appears notably higher expected number of times on no comment ($2.3 > 1.96$), showing that most humanists have no comment about the assistance of TGAZ in

Table VIII. Independent-samples *t*-test of perceived ease of use of technology acceptance of two systems

System	Number of users	Mean	SD	*t*	Sig. (two-tail)
ATAS	31	34.29	4.383	2.038	0.046
MARKUS	31	31.39	6.611		

Table IX. Descriptive statistics of the assistance of linked databases used in the ATAS

Linked source database	Number of users	Mean	SD
Moedict	31	4.06	0.727
Wikipedia	31	3.63	0.809
CBDB	31	3.58	0.886
TGAZ	31	3.48	0.926
EC dictionary	31	2.23	1.055

Table X. One-way analysis of variance of linked databases of the ATAS developed in this study

System	Sum of square	Degree of freedom	Sum of square mean	*F*	Sig.	*Post hoc*
Between groups	59.290	4	14.823	18.789	0.000	Wikipedia > EC dictionary; Moedict > EC dictionary; CBDB > EC dictionary; TGAZ > EC dictionary
Intragroup	117.547	149	0.789			

	Extremely helpless	Helpless	No comments	Helpful	Extremely helpful
Wikipedia					
Number	1	3	8	16	3
Expected number	2.2	4.2	8.0	12.8	3.8
Residual	−1.2	−1.2	0.0	3.2	−0.8
Adjusted residual	−0.9	−0.7	0.0	1.3	−0.5
Moedict					
Number	0	1	4	18	8
Expected number	2.2	4.2	8.0	12.8	3.8
Residual	−2.2	−3.2	−4.0	5.2	4.2
Adjusted residual	−1.7	−1.9	−1.8	2.1	2.6
CBDB					
Number	1	2	9	16	3
Expected number	2.2	4.2	8.0	12.8	3.8
Residual	−1.2	−2.2	1.0	3.2	−0.8
Adjusted residual	−0.9	−1.3	0.5	1.3	−0.5
TGAZ					
Number	1	2	13	11	4
Expected number	2.2	4.2	8.0	12.8	3.8
Residual	−1.2	−2.2	5.0	−1.8	0.2
Adjusted residual	−0.9	−1.3	2.3	−0.7	0.1
EC dictionary					
Number	8	13	6	3	1
Expected number	2.2	4.2	8.0	12.8	3.8
Residual	5.8	8.8	−2.0	−9.8	−2.8
Adjusted residual	4.5	5.2	−0.9	−4.0	−1.7

Table XI. Pearson's χ^2 test of linked databases of the ATAS developed in this study

the text interpretation. EC dictionary shows significantly higher expected number of times on extremely helpless ($4.5 > 1.96$) and helpless ($5.2 > 1.96$), and notably lower expected number of times on helpful ($-4.0 < -1.96$). It reveals that most research participants are inclined to the helplessness of EC dictionary in the text interpretation.

5.5 Analysis of use behavior of ATAS

To analyze the behavioral transfer of the research participants who used the ATAS, the system's operation behaviors of the research participants were encoded with time sequence for a series of behavior sequence samples, according to the system's functions, for lag sequential analysis. To perform the lag sequential analysis, the number of samples in sequential analyses was calculated by frequency of the neighboring pairs of events. The zero-order model proposed by Bakeman (1986) was used to calculate the Z score. The calculation is suitable for samples with a non-normal distribution when the probability of sequence is equal. A Z score above 1.96 indicates that the sequence presents remarkable coding transfer that the research participants with obvious behavioral transfer in the system's operation could be observed, and a high Z score indicates a larger behavioral transfer compared to a low Z score. Figure 8 shows the behavior transfer of the research participants who used the ATAS to support their understanding while reading a text of Ming dynasty's collections.

In Figure 8, 1~6 stand for various use behaviors of the ATAS, and the arrows between use behaviors represent the behavior transfer; the thicker line stands for more obvious behavior transfer. The value on the arrow is Z mark, representing the significance of

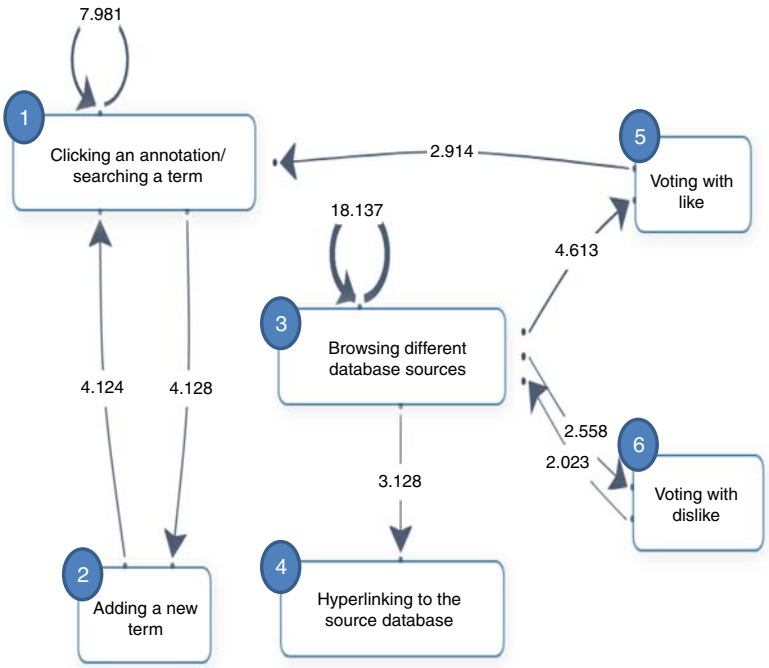

Figure 8.
Behavior transfer of ATAS

behavior transfer. The significant level is achieved when the value is larger than 1.96, revealing the remarkably large number of behavior sequence. The following sequential behavior transfer is explained as below:

(1) Clicking an annotation/searching a term: repeating such behavior sequence to achieve the significant level ($Z = 7.981 > 1.96$) reveals that there are many users constantly clicking the annotation for data search or searching terms without annotation.

(2) Adding a new term: adding a new term after the term search would have the behavior transfer reach the significant level ($Z = 4.128 > 1.96$), revealing that a lot of users would add terms without annotation in the system as new terms. The behavior transfer of browsing other annotations also achieves the significance ($Z = 4.124 > 1.96$).

(3) Browsing different database sources: the behavior of continuously clicking on different databases sources for browsing achieves the significant level ($Z = 18.137 > 1.96$), showing that most users would refer to different linked databases in the ATAS.

(4) Hyperlinking to the source database: after browsing different data sources, the behavior transfer of linking to the original websites of source database reaches the significant level ($Z = 3.128 > 1.96$), showing that the users would further link to the original websites for browsing. Such a result is consistent with the respondents' answers in the interview analyses.

(5) Voting with like: after browsing different data sources, the behavior transfer of clicking on the Helpful button achieves the significant level ($Z = 4.613 > 1.96$), and

the behavior transfer of continuously viewing other annotated terms after clicking on the Helpful button also reaches the significant level ($Z = 2.914 > 1.96$). It shows that most users would not return to see the sources of other data, but continuously browsed the article, after helpful information was searched.

(6) Voting with dislike: after browsing different data sources, the behavior transfer of clicking on the Helpless button reaches the significant level ($Z = 2.558 > 1.96$), and the behavior transfer of viewing the other data sources of the same term, after clicking on the Helpless button, also achieves the significance ($Z = 2.023 > 1.96$). Such a result reveals that users, when not finding out helpful information, would return to view the data from other linked databases for other useful information.

5.6 Summary of interviews

To understand the participants' perception when interpreting the texts with the support of ATAS and MARKUS semi-ATAS, the semi-structured interviews were preceded with the 31 research participants. The results are summarized as follows:

(1) Different opinions about the assistance of the automatic segmentation of Chinese word in text interpretation: 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 texts and accelerating the reading. Other research participants considered that the automatic Chinese word segmentation has not much assistance, as they could segment the terms searched by the system or uncertain correctness of Chinese word segmentation might affect the text interpretation.

(2) Linking to source websites for getting more information: 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.

(3) Text segmentation being more important than word segmentation: 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 collections. They also regarded the better importance of text segmentation than word segmentation to smoothen the reading.

(4) Opposite opinions about the benefits of annotation classification in MARKUS to text reading effectiveness: positive opinions appeared on the annotation classification in MARKUS, revealing that the annotations with different colors allow them clearly viewing the different types of terms, e.g. names of people, location and position, so that the reading process becomes clearer. The research participants with negative opinions regarded that the term classification is unnecessary; instead, classifying with part of speech might be more helpful. Besides, terms annotated with different colors would affect reading.

(5) Increase of variations dictionary or variations proofreading: a lot of research participants expressed that there are many variations in 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.

(6) Increasing more LD to make up inadequate data sources: most of the research participants considered that the ATAS has inadequate LD sources. They suggested

increasing databases or dictionaries which are often used for interpreting texts or research, e.g. Chinese dictionary (http://dict.revised.moe.edu.tw/cbdic/), Scripta Sinica Database (http://hanchi.ihp.sinica.edu.tw/ihp/hanji.htm) and Zdic (www.zdic.net/), for expanding the LD sources.

In summary, most of the research participants agreed with the accuracy of automatic annotation based on Chinese word segmentation with manually adding new terms in interpreting the contents of the Chinese ancient texts and accelerating the reading, whereas the negative opinions from the research participants appeared on the MARKUS semi-ATAS, indicating that the term classification is unnecessary as well as the annotated terms with different colors would affect reading. This study logically inferred that these reasons can explain why the proposed ATAS was better accepted and had better outcomes of text summaries than the MARKUS semi-ATAS.

6. Conclusions and future works

By examining the experimental results and the interview from the research participants, the conclusions and future research directions of this study are summarized as follows. First, the mean of reading abstracts with the ATAS support is higher than it with MARKUS semi-ATAS, but does not achieve the statistical significance. It is encouraged that the technology acceptance degree of the ATAS is significantly higher than that of MARKUS semi-ATAS. Particularly, the functions of the ATAS show more ease of use than MARKUS semi-ATAS. Moedict, among the linked databases of the ATAS, shows the best assistance, while EC dictionary shows the least assistance. Also, the function of automatic Chinese word segmentation cannot achieve the extreme accuracy so that some research participants regarded the function as helpless. Most of the research participants expressed that the user interface of a digital humanities research platform is a key factor affecting most humanists to accept the system. Moreover, the function of automatic text segmentation is more important than automatic word segmentation.

Several research topics are proposed for further study. First, comparison of reading the comprehension effectiveness between the ATAS for supporting digital reading and traditional paper-based reading is a valuable research issue. Currently, the ATAS simply supports digital humanities research. In the future, text contents could be changed for digital reading to compare the reading comprehension effectiveness between digital reading with the ATAS support and traditional paper-based reading. Second, mining social networks relationships among characters appearing in the texts by using the ATAS and LD is a potential research issue. The CBDB linked with the ATAS records the biographical data of all important characters in China history, including names of people, nicknames and kinship. Based on such data, mining the social networks relationships among characters in the texts might be able to find out the characters' relationships of which was hardly aware by manpower. Third, developing a dashboard system for real-time analysis of humanists' behavior processes should be considered in the future. The behavior process records of participants using the ATAS could be simply used for *post hoc* analyses, but not giving real-time feedback during users interpreting texts. Aiming at the real-time analysis of the behavior process of humanists interpreting texts, the development of a real-time humanist behavior process analysis dashboard system could provide humanists with more effective information for the text interpretation. Fourth, developing a Chinese parser for Chinese ancient texts is an urgent issue. The Jieba Chinese parser was applied to the ATAS for automatic word segmentation. However, the Jieba Chinese parser is not developed for ancient texts so that the word segmentation of ancient texts could not present high accuracy as for modern texts. Although this system offers the function of adding new terms, it is expected to enhance the accuracy of Chinese word segmentation through semi-automatic

adjustment of the human–computer interaction in the future. Finally, evaluating humanists' digital reading habits and research requirements to improve the system reader also needs to be considered. The difference between digital humanities research and traditional humanities research lies in the change of reading media. To let humanists completely accept the ATAS, it is suggested to precede deeper evaluation and discussion of humanists' digital reading habits and research requirements and continuously improve the reader of the ATAS in the future.

References

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R. and Ives, Z. (2007), "DBpedia: a nucleus for a web of open data", in Aberer, K. *et al.* (Eds), *Lecture Notes in Computer Science*, Vol. 4825, Springer, Berlin and Heidelberg, The Semantic Web, ISWC 2007, ASWC 2007, pp. 722-735.

Bakeman, R. (1986), *Observing Interaction: An Introduction to Sequential Analysis*, Cambridge University Press, New York, NY.

Bennett, R., Hengel-Dittrich, C., O'Neill, E.T. and Tillett, B.B. (2006), "VIAF (virtual international authority file): linking die deutsche bibliothek and library of congress name authority files", *World Library and Information Congress: 72nd IFLA General Conference and Council, Seoul, August 20-24, 2006*.

Berners-Lee, T. and Fischetti, M. (1999), *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*, HarperSanFrancisco, San Francisco, CA.

Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J. and Sheets, D. (2006), "Tabulator: exploring and analyzing linked data on the semantic web", *Proceedings of the 3rd International Semantic Web User Interaction Workshop*, *Athens, GA, November 6, 2006*.

Bizer, C., Heath, T. and Berners-Lee, T. (2009), "Linked data – the story so far", *Journal on Semantic Web and Information Systems*, Vol. 5 No. 3, pp. 1-22.

Chen, C.M. and 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, S.P., Hsiang, J., Tu, H.C. and Wu, M.C. (2007), "On building a full-text digital library of historical documents", *Lecture Notes in Computer Science: Proceedings of the 10th International Conference on Asian Digital Libraries: Looking Back 10 Years and Forging New Frontiers*, *Hanoi*, Vol. 4822, pp. 49-60.

Cole, T.W., Han, M.-J., Weathers, W.F. and Joyner, E. (2013), "Library marc records into linked open data: challenges and opportunities", *Journal of Library Metadata*, Vol. 13 Nos 2-3, pp. 163-196.

Dutta, B. (2017), "Examining the interrelatedness between ontologies and linked data", *Library Hi Tech*, Vol. 35 No. 2, pp. 312-331.

Ho, H.I.B. and Hilde, D.W. (2014), "MARKUS – text analysis and reading platform", available at: http://dh.chinese-empires.eu/markus/beta/ (accessed September 26, 2017).

Hsiang, J., Chen, S.P. and Tu, H.C. (2009), "On building a full-text digital library of land deeds of Taiwan", *Digital Humanities, University of Maryland, College Park, MD, June 22-25, 2009*, pp. 85-90.

Hwang, G.J., Yang, L.H. and Wang, S.Y. (2013), "A concept map-embedded educational computer game for improving students' learning performance in natural science courses", *Computers & Education*, Vol. 69, pp. 121-130.

Kobilarov, G., Scott, T., Raimond, Y., Oliver, S., Sizemore, C., Smethurst, M. and Lee, R. (2009), "Media meets semantic web – how the BBC uses DBpedia and linked data to make connections", *European Semantic Web Conference*, pp. 723-737.

Monte, L.D. and Serafin, M. (2017), "Drupal, TEI and XML: how to prototype a digital humanities tool?", *Library Hi Tech News*, Vol. 34 No. 4, pp. 9-15.

Rosenzweig, R. (2003), "Scarcity or abundance? Preserving the past in a digital era", *The American Historical Review*, Vol. 108 No. 3, pp. 735-762.

Scheinfeldt, T. (2010), "Why digital humanities is 'Nice' ", available at: https://foundhistory.org/2010/05/why-digital-humanities-is-nice/ (accessed September 26, 2017).

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

Tu, A., Hung, J.J. and Lin, Y.H. (2012), "Building a text analysis platform for Chinese Buddhist text – an example based on CBETA and Tripitaka catalog projects", *PNC 2012 Annual Conference and Joint Meetings, UC Berkeley School of Information, Berkeley, December 7-9, 2012*.

About the authors

Chih-Ming Chen is currently a Distinguished Professor in the Graduate Institute of Library, Information and Archival Studies at National Chengchi University, Taipei, Taiwan. He received his BSc and MSc Degrees from the Department of Industrial Education at National Taiwan Normal University in 1992 and 1997, and received his PhD Degree from the Department of Electronic Engineering at National Taiwan University of Science and Technology in 2002. His research interests include digital library, e-learning, data mining, machine learning and intelligent agents on the web. Chih-Ming Chen is the corresponding author and can be contacted at: chencm@nccu.edu.tw

Yung-Ting Chen received his BSc Degree from the Department of Information and Library Science at Fu Jen Catholic University in 2007, and he is now a PhD Candidate of the Graduate Institute of Library, Information and Archival Studies at National Chengchi University. His research interests include digital library and e-learning.

Chen-Yu Liu received his BSc Degree from the Department of Information management at National Chung Cheng University in 2015, and he is now a Graduate Student of the Graduate Institute of Library, Information and Archival Studies at National Chengchi University. His research interests include digital humanities and e-learning.