

The "Database of Full Text Chinese Documents"
at the Institute of History and Philology, Academia Sinica
Jender Lee
Institute of History and Philology, Academia Sinica

Summary

The "Database of Full Text Chinese Documents" (hereafter "Full Text") is the general title of the computerized database of classical Chinese texts developed by the Institute of History and Philology over the past thirteen years. The best-known part of this database is the "Database of the Full Text of the Twenty-five Dynastic Histories" (or "the Twenty-five Dynastic Histories Database"). I will focus on two aspects of the database: the development and current state of the "Full Text," and its application for historical and linguistic research. To my knowledge there are so far at least four articles in Chinese that introduce the Full Text.¹ This summary relies heavily on these articles. Readers interested in more information about the database are encouraged to read these articles.

1. The Development and Current State of the "Full Text"

The original project began under the name "Computerization of Historical Documents" in 1984 when researchers from both the Institute of History and Philology and the Computing Center at Academia Sinica worked together to key in the "Monographs on Economy" from the dynastic histories. In 1986, the project was expanded to include the entire twenty-five dynastic histories. In June 1990 the computerization of the full text of the dynastic histories was completed with the exception of the "charts." This part of the project took six years and cost more than NT\$40,000,000 (approximately US\$1,400,000). The "database of the twenty-five dynastic histories" is the first and largest segment of the Full Text Project.

¹ They are, 1) Huang Ch'ing-lien, "The 'Database of the Full Text of the Twenty-five Dynastic Histories' as a Research Tool for Chinese History," *New History* 2:2(1991); 2) Liu Tseng-kuei, "Introduction to the 'Database of the Full Text of the Twenty-five Dynastic Histories'", *Newsletter of the Central Advisory Committee of Academia Sinica* 2:1(1993); 3) Hsieh Ch'ing-chun & Lin Hsi, "On the 'Database of the Full Text of Classical Documents'", *Newsletter of the Computer Center of Academia Sinica* 11:15-17(1995); 4) Lee Jender & Ch'en Jo-shui, "On the 'Database of the Full Text of Chinese Documents' by the Institute of History and Philology," *Newsletter of the Chinese Society of Library Sciences* 4:3 (1996). Readers are also encouraged to consult *The Full-Text Retrieval System-Users Manual* by Mr. Lin Shih, published by the Computing Center, Academia Sinica (translation reviewer: Karl K. Lo)

After the completion of the computerization of the dynastic histories, the project was expanded to include the "Database of the Thirteen Classics and their Annotations" and other classical Chinese texts. It was also renamed the "Database of Full Text Chinese Documents." By August 1996, this project had completed 69 databases, including approximately 71,420,000 Chinese characters. There are 12 other databases involving 50,610,000 Chinese characters currently under construction. On completion, the Full Text Database will include more than 122,030,000 Chinese characters, and will likely be the largest database of computerized Chinese texts in the world. (See appendix)²

Most texts before the T'ang dynasty (618-907) are now computerized and available through either purchase or registration.³ The current key-in work is focused on encyclopedia texts from the Sung dynasty (960-1279) to the present time. The software used in the Full Text Project, CTP 中文全文檢索系統, is part of the Data Base Management System (DBMS) developed by Academia Sinica. In an earlier stage, the Institute of History and Philology expended tremendous human resources for key-in, proof-reading, revision and mark-up (i.e., writing the CTP into the plain text). Now the process has been greatly simplified. The key-in work is basically done by professionals who hold contracts with the Institute. Each document is typed by two different groups before being proof-read both manually and by computer program. The work done in the Institute is now mainly mark-up.

But the project has its problems and is under revisions from time to time. The first problem is that the selected texts have different versions or sometimes belong to different collections. Now we have included the "Pre-Ch'in and Western Han Chinese Corpus" 先秦兩漢語料庫 in the Full-Text Database. The "Pre-Ch'in and Western Han Chinese Corpus" is developed by the Section of Linguistics in the Institute of History and Philology. The Corpus includes almost all texts 子書 (annotation not included) of the Pre-Ch'in and Western Han period and is very convenient for search. The second problem is that names used for grouping texts are overlapping or ambiguous. For example, names such as 古籍十八種、古籍十九種、古籍三十四種(18 Classical Texts, 19 Classical Texts, 34 Classical Texts, etc.) need to be regrouped or renamed.

2 As can be seen from the Appendix that at the beginning of this year (1997) nearly 80 millions (79,137,416) Chinese characters are already available for search.

3 In general, price for each CD-ROM varies according to the amount of Chinese characters included in that specific database. Take the twenty-five dynastic histories for example, it includes nearly 40 million Chinese characters, and it now costs 30 thousand US dollars. Registration account will be open on March 1. Annual registration fee for each machine is NT\$8,000 (app. US\$290). Registration for one Class C network(256 PC):NT\$5,000/year(US\$180). Once registered, users can access to all databases on line. Registration open only to users in Taiwan.

But there are some recent progress also. First, the "charts" in the Twenty-five dynastic histories were not included in the first stage of development due to technical difficulties. Now the difficulties are overcome; "charts" have been key-in and are now under proof-reading. Second, in order to better utilize the Full-Text project, cooperation among institutions is promoted. For instance, the Institute of History and Philology is now working with the Hong Kong Chinese University to establish database of Neo-Confucian texts.

After more than ten years of hard work, the Full Text Project is now very convenient to access, and has already proven itself a valuable tool for scholars of Chinese studies.

2. General Applications of the "Full Text"

Since the Database is Full Text, it provides at least three functions: 1) Free reading. Users can read texts on screen either by following the directory and page number or by reading selections according to their own needs. 2) Free text search. Users can type any character or term that they wish to find, and search the whole text without limitation on the amount of character/term. 3) Printing and saving. Users can print/save either the page number, the sentence, or the paragraph of the character/term they have located with the search. The newly established WWW search system is even more accessible than the DOS version.

Since the completion of the Database of the Full Text of the Twenty-five Dynastic Histories in 1990, scholars in the Institute of History and Philology have nearly six years experience in using it. Different academic institutes and universities that have purchased the CD-ROM have also shared their experience with us. Scholars who have used the Full Text agree that the most important aspects of the Full Text are its speed and comprehensiveness. Traditionally when historians collected material for research, they would carefully read through often lengthy documents, write down related sentences/paragraphs, and then copy these into their drafts again when they wrote up their articles. This process takes a tremendous amount of time and runs the risk of accidentally missing an important piece of evidence. Now with the Full Text, scholars are able to search through huge amounts of documents in several seconds without fear of overlooking an important entry.

Take, for example, the system set up on UNIX in the Institute of History and Philology. It takes only one or two seconds to search one character/term throughout the twenty-five dynastic histories, a collection containing almost 40,000,000 characters in total. Users can read the documents on screen and print/save useful material after primary selection.

In application, since different parts of the same document or several different documents can be consulted at the same time during the search event, the Full-Text is very useful in collecting lost texts or material that are cited in accessible documents. Two examples are

given below. First, when P'ei Sung-chih(裴松之) wrote his commentary on the *History of the Three Kingdoms*, he often quoted *Wei Lueh* (魏略). Since *Wei Lueh* is now lost, scholars who want to gather together the text can search either Sung-chih or *Wei Lueh* in the *History of the Three Kingdoms* to find all the cited documents. Second, when the Han dynasty scholars Cheng Hsuan(鄭玄)and Cheng Chung(鄭眾)wrote their commentaries on the Classics, they often cited Han law or Han custom to give examples. Since there is no complete law book from the Han dynasty available now, their commentaries are useful material. Scholars who want to study the Han law can search the character/term "Han/Han Lu (漢/漢律) in Classics to gather together Han statutes.

The powerful search function of the Full-Text Project also encourages new research subjects that have traditionally been hindered by the overwhelming quantity of the material involved. For instance, women's history, of which the material is scattered in documents; medical history and religious history that have enormous amount of material but not easily tangible. They are all made possible by using the Full-text Project to do some primary search. The decision in selecting documents for computerization also involves the creation of new research.

However, the Full Text has its pros as well as its cons. Since it can be searched freely with all characters and terms, users have to establish basic parameters for their research subjects before using the database in order to avoid too many unwanted sentences/paragraphs on screen. The example used in chapter 4 of the *Full-Text Retrieval System-Users Manual* is "feng"(風), literary "wind" in Chinese. Users who search for the character 風 to study weather could get numerous other character sets which contain the word 風 but have nothing to do with weather, such as "feng-shang"(風尚 fashion), "feng-ssu"(風俗 custom), "feng-ko"(風格 manner or style) etc. The *Manual* suggests several solutions. But in the end, as in the case of most new research tools, the computerization of Chinese texts is most helpful when scholars not only use the technology to replace traditional methods of research, but also explore ways of exploiting the technology for new and innovative forms of research.

中央研究院歷史語言研究所「漢籍全文資料庫工作室」工作概況

1997.2.5

「古籍三十四種」、「大正新脩大藏經」

1. 「廿五史全文資料庫」總目錄及頁數：(台灣鼎文書局)

【代碼】	【內 容】	【頁 碼】
1.	史籍自動化的資料校對說明	1- 13
2.	新校本史記三家注	
3.	新校本漢書	

4.	新校本後漢書	1- 3684
5.	新校本三國志	1- 1510
6.	新校本晉書	1- 3216
7.	新校本宋書	1- 2471
8.	新校本南齊書	1- 1036
9.	新校本梁書	1- 867
10.	新校本陳書	1- 499
11.	新校本魏書	1- 3062
12.	新校本北齊書	1- 698
13.	新校本周書	1- 930
14.	新校本南史	1- 2027
15.	新校本北史	1- 3351
16.	新校本隋書	1- 1904
17.	新校本舊唐書	1- 5407
18.	新校本新唐書	1- 6472
19.	新校本舊五代史	1- 2042
20.	新校本新五代史	1- 923
21.	新校本宋史	1-14263
22.	新校本遼史	1- 1560
23.	新校本金史	1- 2906
24.	新校本元史	1- 4678
25.	新校本明史	1- 8642
26.	新校本清史稿	1-14740

以上可供查詢之總字數共計約 39,311,460 字。

2. 「古籍十九種」總目錄及其內容、頁數如下：

【代碼】	【內 容】	【頁 碼】
1.	抱朴子(抱朴子內篇校釋)(王明, 中華; 台北里仁, 1991)	1- 363
2.	莊子集釋([清]郭慶藩, 中華; 台北華正, 1987)	1- 106
3.	法言義疏(汪榮寶, 北京中華, 1987)	1- 628
4.	東觀漢記(東觀漢記校注) ([東漢]劉珍等, 河南中州古籍, 1986)	1- 934
5.	墨子城守各篇簡注(岑仲勉, 中華, 1959)	1- 155
6.	潛夫論箋(潛夫論箋校正) ([清]汪繼培中華, 1985; 台北大立, 1984)	1- 492
7.	國語([春秋]左丘明, 中華; 台北里仁/宏業, 1980)	1- 659
8.	莊子集解([清]王先謙, 中華, 1987; 台北文津, 1988)	1-
299		
9.	莊子集解內篇補正([清]劉武, 中華, 1987; 台北文津, 1988)	1-
198		
10.	古本竹書紀年輯證(方詩銘/王修齡, 上海古籍, 1981)	1- 294
11.	墨子閒詁([清]孫詒讓, 中華; 台北華正, 1987)	1-

711	12. 列子集釋(楊伯峻, 中華, 1979; 台北華正, 1987)	1-
350	13. 晏子春秋集釋(吳則虞, 中華; 台北鼎文, 1977)	1-
667	14. 管子輕重篇新詮(馬非白, 中華, 1979)	1-
748	15. 點校四書章句集注([宋]朱熹, 中華, 1986; 台北大安, 1987)	1-
388	16. 新語校注(王利器, 中華, 1986)	1-
252	17. 戰國策([西漢]劉向, 上海古籍, 1978)	1-
1366	18. 八家後漢書(八家後漢書輯注)(周天游, 上海古籍, 1986)	1-
802	19. 老子校釋(朱謙之, 中華; 台北華正, 1986)	1-
340		

以上可供查詢之字數共計約 5,856,530 字。

3. 「十三經」:

「十三經注疏」總目錄及其內容、頁數如下:(阮刻本)

【代碼】	【內 容】	【頁 碼】
1.	周易	1- 215
2.	尚書	1- 318
3.	毛詩正義	1- 809
4.	周禮注疏	1- 668
5.	儀禮注疏	1- 611
6.	禮記注疏	1- 1037
7.	春秋左傳正義	1- 1065
8.	春秋公羊傳注疏	1- 361
9.	春秋穀梁傳注疏	1- 207
10.	論語注疏	1- 181
11.	孝經注疏	1- 58
12.	爾雅注疏	1- 206
13.	孟子注疏	1- 268

以上可供查詢之字數共計約 7,928,722 字。

「斷句十三經經文」總目錄及其內容、頁數如下:(台北開明書局)

【代碼】	【內 容】	【頁 碼】
1.	周易	1- 31
2.	尚書	1- 41
3.	毛詩	1- 89

4.	周禮	1-	75
5.	儀禮	1-	76
6.	禮記	1-	134
7.	春秋左傳	1-	272
8.	春秋公羊傳	1-	67
9.	春秋穀梁傳	1-	62
10.	論語	1-	23
11.	孝經	1-	4
12.	爾雅	1-	15
13.	孟子	1-	48

以上可供查詢之字數共計約 730,000 字。

4. 「古籍十八種」總目錄及其內容、頁數如下：

【代碼】	【內 容】	【頁 碼】
1.	唐令拾遺(栗勁等譯, 長春出版社)	1- 926
2.	新校搜神記(世界書局本)	1- 156
3.	齊民要術校釋(繆啓愉校釋, 農業出版社)	1- 870
4.	世說新語箋疏(余嘉錫)	1- 933
5.	典論(世界書局本)	1- 11
6.	申鑒(世界書局本)	1- 29
7.	中論(世界書局本)	1- 42
8.	漢官六種(孫星衍等輯, 周天游點校, 中華)	1- 220
9.	洛陽迦藍記校注(范祥雍校注)	1- 387
10.	九家舊晉書輯本(湯球輯, 鼎文版)	1- 544
11.	顏氏家訓集解(顏之推撰, 王利器集解, 上海古籍出版社)	1- 642
12.	荆楚歲時記(守屋美都雄, 中國古歲時記的研究)	1- 72
13.	唐律疏議(劉俊文點校, 北京中華書局)	1- 678
14.	山海經校注(袁珂點校, 上海古籍, 1980)	1- 495
15.	通典(中華書局點校本)	1- 5646
16.	風俗通義校注(應邵撰, 王利器校注, 北京中華, 1981)	1- 652
17.	唐會要(世界書局本, 影上海中華書局)	1- 1799
18.	後漢紀校注(周天游, 天津古籍出版社)	1- 911

以上可供查詢之總字數共計約 8,049,602 字

5. 「古籍三十四種」總目錄及其內容、頁數如下：

【代碼】	【內 容】	【頁 碼】
1.	鄧析子(新編諸子集成六, [周]鄧析, 世界書局)	1- 31
2.	關尹子(四部備要, 台灣中華書局, 據金壺本校刊)	1- 51
3.	太平經合校(王明, 北京, 中華書局, 1960)	1- 763
4.	鬼谷子(陶弘景注, 台灣商務)	1- 72
5.	尹文子(新編諸子集成六, [清]錢熙祚校, 世界)	1- 16
6.	慎子(新編諸子集成五, [清]錢熙祚校, 世界)	1- 15

7.	孔子家語(新編諸子集成二, [魏]王肅注, 世界書局)	1- 115
8.	鶡冠子(四部備要, 台灣中華)	1- 132
9.	通玄真經(文子)(四部叢刊)	1- 214
10.	孔叢子(增訂漢魏叢書(三), [清]王謨輯, 大化書局)	
1581-1618		
11.	藝文類聚(木鐸, 標點本)	1- 1733
12.	論衡校釋(黃暉, 北京中華書局, 諸子集成本)	1- 1366
13.	金匱要略(李克光, 知音出版社)	17- 699
14.	難經本義新解(林輝鎮, 益群)	27-
324		
15.	傷寒論(李培生, 知音出版社)	10-
751		
16.	黃帝內經(楊維傑, 台聯國風出版社)	
17.	前漢紀(國學基本叢書四百種, 荀悅撰, 台灣商務)	1- 313
18.	漢魏南北朝墓誌彙編(趙超, 天津古籍出版社)	1- 510
19.	九章算經點校(錢寶琮, 九章出版社)	1- 232
20.	周髀算經(叢書集成, 上海商務)	1- 94
21.	越絕書(李步嘉, 武漢大學出版社)	1- 340
22.	釋名(增訂漢魏叢書(一), [清]王謨輯, 大化書局)	833- 888
23.	方言校箋(方言校箋附通檢, 周祖謨, 鼎文)	1- 89
24.	穆天子傳(四部叢刊)	1- 63
25.	西京雜記(增訂漢魏叢書(二), [清]王謨輯, 大化書局)	
1071-1092		
26.	吳越春秋(四部叢刊)	1- 277
27.	逸周書(皇清經解續編(三), 漢京文化事業有限公司)	
1924-1994		
28.	文獻通考(馬端臨, 商務)	1- 2764
29.	朱子語類([宋]黎靖德編, 王星賢點校, 世華出版社)	1- 3343
30.	楚辭補注([宋]洪興祖, 天工)	1- 327
31.	敦煌變文集新書(潘重規, 中國文化大學)	1- 1266
32.	文選([梁]蕭統編, [唐]李善注, 文津)	1- 2610
33.	華陽國志校補圖注(任乃強, 上海古籍出版社, 1987)	1- 773
34.	古小說鈎沈(魯迅, 盤庚出版社, 文史叢刊)	1-

539

以上可供查詢之總字數共計約 12,264,713 字

6. 「大正新脩大藏經」內容如下：約 5,054,793 字

【 內 容 】

迦葉摩騰共法蘭： 784.四十二章經

安世高：

13.長阿含十報法經

14.佛說人本欲生經

- | | |
|----------------------|---------------|
| 31.佛說一切流攝守因經 | 32.佛說四諦經 |
| 36.佛說本相猗致經 | 48.佛說是法非經 |
| 57.佛說漏分布經 | 98.佛說普法義經 |
| 91.佛說婆羅門子命終愛念不離經 | 92.佛說十支居士八城人經 |
| 105.五陰譬喻經 | |
| 109.佛說轉法輪經 | 112.佛說八正道經 |
| 131.佛說婆羅門避死經 | 140.阿那邠邸化七子經 |
| 149.佛說阿難同學經 | 151.佛說阿含正行經 |
| 150A.佛說七處三觀經 | 150B.佛說九橫經 |
| 167.佛說太子慕魄經 | 348.佛說大乘方等要慧經 |
| 356.佛說寶積三昧文殊師利菩薩問法身經 | |
| 492.佛說阿難問事佛吉凶經 | 506.犍陀國王經 |
| 525.佛說長者子懊惱三處經 | 526.佛說長者子制經 |
| 551.佛說摩登女解形中六事經 | 553.佛說 女祇域因緣經 |
| 554.佛說柰女耆婆經 | 604.佛說禪行三十七品經 |
| 603.陰持入經 | 602.佛說大安般守意經 |
| 605.禪行法想經 | 607.道地經 |
| 621.佛說佛印三昧經 | 622.佛說自誓三昧經 |
| 684.佛說父母恩難報經 | 701.佛說溫室洗浴眾僧經 |
| 724.佛說罪業應報教化地獄經 | 729.佛說分別善惡所起經 |
| 730.佛說處處經 | 731.佛說十八泥犁經 |
| 732.佛說罵意經 | 733.佛說堅意經 |
| 735.佛說鬼問目連經 | 779.佛說八大人覺經 |
| 1467.佛說犯戒罪報輕重經 | 1470.大比丘三千威儀 |
| 1492.佛說舍利弗悔過經 | 1557.阿毘曇五法行經 |
| 2027.迦葉結經 | |

支婁迦讖：

- | | |
|----------------|-------------------|
| 204.雜譬喻經 | |
| 224.道行般若經 | 280.佛說兜沙經 |
| 313.阿門佛國經 | 350.佛說遺日摩尼寶經 |
| 361.佛說無量清淨平等覺經 | |
| 417.佛說般舟三昧經 | 418.般舟三昧經 |
| 458.文殊師利問菩薩署經 | 624.佛說佉真陀羅所問如來三昧經 |
| 626.佛說阿闍世王經 | 807.佛說內藏百寶經 |

嚴佛調： 778.佛說菩薩內習六波羅蜜經

安玄共嚴佛調： 322.法鏡經

1508.阿含口解十二因緣經

支曜：

- | | |
|---------------|-------------|
| 46.佛說阿那律八念經 | 114.佛說馬有三相經 |
| 115.佛說馬有八熊譬人經 | 608.小道地經 |
| 630.佛說成具光明定意經 | |

康孟詳：

137.舍利弗摩訶目連遊四衢經 197.佛說興起行經

曇果共康孟詳：	196.中本起經	
康孟詳共竺大力：	184.修行本起經	
支謙：	198.佛說義足經	225.大明度經
維祇難：	210.法句經	
康僧會：	152.六度集經	
竺法護(晉)：	154.生經	186.佛說普耀經
法矩共法立(晉)：	23.大樓炭經	211.法句譬喻經
竺佛念(姚秦)：	212.出曜經	
鳩摩羅什(後秦)：	208.眾經雜撰譬喻	223.摩訶般若波羅蜜經
曇無讖(北涼)：	157.悲華經	
慧覺(元魏)：	202.賢愚經	
求那毘地(蕭齊)：	209.百喻經	
闍那崛多：	190.佛本行集經	
佛陀耶舍共竺佛念：	1428.四分律	
弗若多羅共羅什：	1435.十誦律	
慧皎(梁)：	2059.高僧傳	
道宣(唐)：	2060.續高僧傳	2103.廣弘明集
僧祐(梁)：	2102.弘明集	

二、進行校對書籍： 10,989,000 字

1.大正藏經：約 4,349,000 字

禪宗語錄等二十三種	No.1985-2006,2012
宋高僧傳	No.2061 [宋]贊寧
比丘尼傳	No.2063 [梁]寶唱等
景德傳燈錄	No.2076 [宋]道原
集古今佛道論衡	No.2104 [唐]道宣
續集古今佛道論衡	No.2105 [唐]智昇
經律異相	No.2121 [梁]寶唱等
法苑珠林	No.2122 [唐]道世
出三藏記集	No.2145 [梁]僧祐

- 2.《水經注》：[後魏]酈道元注，陳橋驛校熊本，江蘇古籍出版社 2,275,000 字
3.《太平御覽》：商務，四部叢刊三編本 4,365,000 字

三、列入比對，第二份尙未完成輸入者：約 36,545,000 字

- 1.《全上古三代秦漢三國六朝文》：[清]嚴可均校輯，中華書局 5,390,000 字
2.續資治通鑑長編(2-34 冊) [宋]李燾，北京中華書局 7,110,000 字
3.明實錄(1-133 冊) 中央研究院歷史語言研究所校勘 2,320,000 字
4.大正藏經 1,500,000 字

	歷代三寶紀	No.2034[隋]費長房	
	佛祖統紀	No.2034 [宋]志磐	
	佛祖歷代通載	No.2036 [元]念常	
	大唐西域求法高僧傳	No.2066 [唐]義淨	
	大唐西域記	No.2087 [唐]玄奘、辯機	
	集神州三寶感通錄	No.2106 [唐]道宣	
字	5.本草綱目	[明]李時珍, 北京人民衛生出版社, 1975	2,180,000
	6.水滸全傳	[元]施耐庵、羅貫中、王利器校訂, 貫雅文化	1,010,000 字
	7.繡像金瓶梅詞話	雪山圖書有限公司(故宮藏萬曆丁巳本)	900,000
字	8.西遊記	[明]吳承恩, 桂冠圖書公司(世德堂本)	865,000 字
	9.紅樓夢校注	[清]曹雪芹、高鶚, 里仁書局(庚辰本、程甲本合配本)	950,000 字
	10.兒女英雄傳	桂冠圖書公司(北京聚珍活字本)	505,000
字	11.儒林外史	[清]吳敬梓, 桂冠圖書公司(臥賢草堂本)	320,000 字
	12.唐語林校證	[宋]王讜撰, 周勛初校證, 北京中華書局, 1987	560,000
字	13.四朝聞見錄	[宋]葉紹翁撰, 沈錫麟、馮惠民點校, 北京中華書局, 1989	138,000 字
	14.大唐新語	[唐]劉肅撰, 許德楠、李鼎霞點校, 北京中華書局, 1984	20,000 字
	15.歸田錄	[宋]歐陽修撰, 李偉國點校, 北京中華書局, 1981	39,000
字	16.泊宅編	[宋]方勺撰, 許沛藻、楊立揚點校, 北京中華書局, 1983	74,000
字	17.清波雜誌校注	[宋]周輝撰, 劉永翔校注, 北京中華書局, 1994	370,000
字	18.湘山野錄, 續湘山野錄	[宋]文瑩撰, 鄭世剛、楊立揚點校, 北京中華書局, 1984	63,000 字
	19.玉壺清話	[宋]文瑩撰, 鄭世剛、楊立揚點校, 北京中華書局, 1984	74,000 字
	20.舊聞證誤	[宋]李心傳撰, 崔文印點校, 北京中華書局, 1981	46,000
字	21.遊宦紀聞	[宋]張世南撰, 張茂鵬點校, 北京中華書局, 1981	58,000
字	22.鐵圍山叢談	[宋]蔡絛撰, 馮惠民、沈錫麟點校, 北京中華書局, 1983	76,000 字
	23.青箱雜記	[宋]吳處厚撰, 李裕民點校, 北京中華書局, 1985	90,000
字	24.鶴林玉露	[宋]羅大經撰, 王瑞來點校, 北京中華書局, 1983	235,000
字	25.齊東野語	[宋]周密撰, 張茂鵬點校, 北京中華書局, 1983	232,000
字	26.龍川別志	[宋]蘇轍撰, 俞宗憲點校, 北京中華書局, 1982	27,000

27.龍川略志	[宋]蘇轍撰，俞宗憲點校，北京中華書局，1982	46,000
字		
28.涑水記聞	[宋]司馬光撰，鄧廣銘、張希清點校，北京中華書局，1989	277,000
字		
29.澗水燕談錄	[宋]王闢之撰，呂友仁點校，北京中華書局，1981	80,000
字		
30.助編	[宋]莊綽撰，蕭魯陽點校，北京中華書局，1983	94,000
字		
31.家世舊聞	[宋]陸游撰，北京中華書局，1993	35,000
字		
32.明皇雜錄	[宋]鄭處誨撰，北京中華書局，1994	36,000 字
33.春明退朝錄三卷	[宋]宋敏求撰，誠剛點校，北京中華書局，1980	34,000
字		
34.東坡志林五卷	[宋]蘇軾撰，王松齡點校，北京中華書局，1981	69,000
字		
35.東觀奏記	[唐]裴庭裕撰，北京中華書局，1994	68,000 字
36.東軒筆錄	[宋]魏泰撰，李裕民點校，北京中華書局，1983	110,000
字		
37.東齋記事五卷	[宋]范鎮撰，汝沛點校，北京中華書局，1980	48,000 字
38.燕翼詒謀錄五卷	[宋]王 撰，誠剛點校，北京中華書局，1981	35,000
字		
39.石林燕語	[宋]葉夢得撰，宇文紹奕考異，侯忠義點校，北京中華書局，1984	14,000 字
40.西溪叢語	[宋]姚寬撰，北京中華書局，1993	79,000 字
41.默記三卷	[宋]王銍撰，朱杰人點校，北京中華書局，1981	34,000 字
42.程史	[宋]岳珂撰，北京中華書局，1981	108,000 字

四、準備輸入部分：**8,140,000** 字

明實錄附錄(3-21 冊)	2,200,000 字
明實錄校勘記(29 冊)	5,940,000 字