

DISCOVERING RELATIONSHIPS FROM IMPERIAL COURT  
DOCUMENTS OF QING CHINA

---

JIEH HSIANG, SHIH-PEI CHEN, HOU-IEONG HO  
AND HSIEH-CHANG TU

***Abstract** The Qing Imperial Court documents are a major source of primary research material for studying the Qing era China since they provide the most direct and first-hand details of how national affairs were handled. However, the way Qing archived these documents has made it cumbersome to collect documents covering the same event and rebuild their original contexts. In this paper, we describe some information technology that we have developed to discover two important and useful relations among these documents. The first is the citation relation among the Imperial Edicts and the Memorials. We discovered 6,801 pairs from the 37,831 Taiwan-related Imperial Court documents in the Taiwan History Digital Library (THDL) and produced 1,101 graphs of successive citations, which we call IE-M diagrams. The second relation is a template relation, which indicates groups of documents that were created following a specific format. Numerical data can also be tabulated from these documents and be used for further analysis. Our studies show how information technology can be used to discover useful contexts from seemingly unrelated historical documents.*

**Keyword:** Qing Dynasty, Digital Libraries, Text Mining, Imperial Court Documents, Citation Relation

---

*International Journal of Humanities and Arts Computing* 6.1–2 (2012): 22–41

Edinburgh University Press

DOI: 10.3366/ijhac.2012.0036

© Edinburgh University Press and the Association for History and Computing 2012

[www.eupjournals.com/ijhac](http://www.eupjournals.com/ijhac)

INTRODUCTION

THDL, the Taiwan History Digital Library, is a full-text digital library of primary historical documents about Taiwan.<sup>1</sup> One of its collections consists of Taiwan-related Imperial Court documents from the Ming and Qing dynasties. The documents, currently numbered at 37,831 and containing 23,000,000 words, are extracted from 235 sources including books and unpublished archives<sup>2</sup>, and are mostly Memorials and Imperial Edicts. They contain a tremendous wealth of information on how the Qing central government dealt with events occurring in Taiwan, a far-flung island of the vast empire.

The large quantity of documents and their diverse origins poses a problem for users. That is, unlike conventional archives whose contents are organised in a certain way, the documents in THDL, being extracted from different sources, no longer possess their original context. In order to help users find collective meanings among these documents, THDL has developed some novel features. One of them is post-classification of query results. THDL classifies the query return according to attributes such as year, source, author, and the nature of documents. Co-occurrence analysis of terms provides another way for the user to observe relations among documents resulting from a query.<sup>3</sup>

This paper describes two important relations among documents in THDL that we have explored through information technology. These relations provide the documents with new contexts that would be very difficult to find without the help of IT. The first is a *citation relation* among the Imperial Edicts and the Memorials: 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. The algorithms we developed discovered 6,801 pairs of citation relations, among which 4,883 pairs are Memorials citing Edicts and 1,918 pairs are Edicts citing Memorials. By taking the transitivity of the citation pairs we have discovered large graphs of successive citations, which we call *IE-M diagrams*. Our discovery produced 1,101 such diagrams, the largest of which is about the Lin Shuangwen Rebellion and involves 153 such documents.

The second relation is a *template* relation, which indicates groups of documents that were created following a specific format. The largest group we found consists of 221 Memorials reporting the prices of grains in the major prefectures of the Fujian Province. Numerical data contained in these documents can also be extracted and tabulated for further analysis.

Our studies show how information technology can be used to discover useful relations from seemingly unrelated historical documents. We hope that they are useful to historians who want to use THDL in their research.

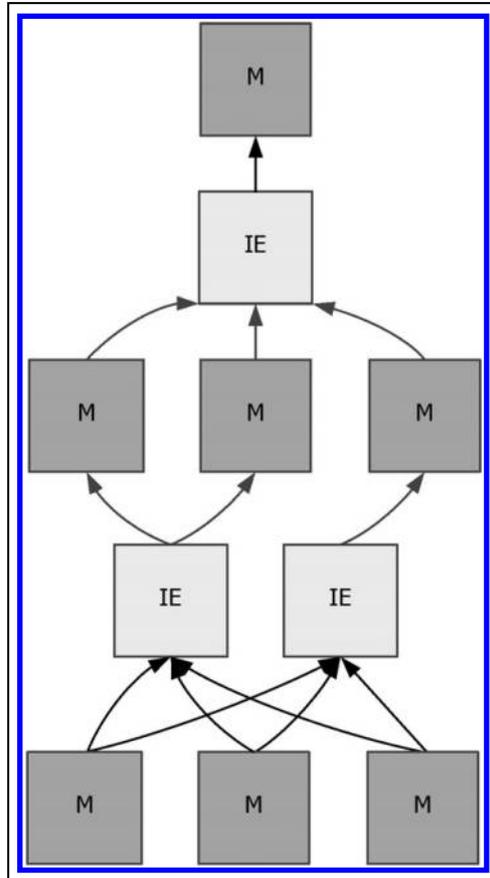
DISCOVERING THE CITATION RELATIONS

2.1 Motivation

Each Chinese dynasty had its own sophisticated governing system<sup>4</sup>. Being the latest and most recent, the Qing Dynasty (1644–1911) also left the largest volume of court documents that are invaluable primary material for the study of how Qing conducted the national affairs. Among them, two of the most important types are *Imperial Edicts* (from the emperors to his officials) and *Memorials* (reports from officials to the emperor). The number of Memorials increased significantly after Emperor Kangxi (康熙, 1662–1723) allowed senior *local* officials to report to him directly<sup>5</sup>. The ability for the emperors to obtain first-hand information directly from local officials was among the major reasons why the Qing imperial courts did not suffer as much interference from people surrounding the emperors, such as eunuchs and family members of the empress dowagers, as in the previous Ming Dynasty.

The court documents of Qing were kept in different offices and were organised chronologically. Although this way of organisation is convenient for record keeping, it created problems for researchers who are interested in studying specific events, since related materials may span over months and be kept in different archives. For instance, if the emperor received a Memorial reporting a rebellion in some province, he might decide to issue an Imperial Edict to give instructions to relevant officials. The Memorial, depending on its nature or origin, might be kept (or had copies made) in the Archives of the Imperial Palace (宮中檔), Archives of the Grand Council (軍機處檔案), or the Grand Secretariat Archives (內閣大庫). The Imperial Edict might have records in the Imperial Decrees Archives (上諭檔), Archives of the Diary-Keeper (月摺檔, 起居注), the Archives of the Imperial Palace, or the Grand Secretariat Archives mentioned above. Worse yet, these archives are now being kept at different geographic locations, notably the National Palace Museum<sup>6</sup>, the Institute of History and Philology of Academia Sinica<sup>7</sup>, both in Taipei, and the First Historical Archives of China<sup>8</sup>, in Beijing. Although digitisation effort at the former two institutions has made their archives easier to access than before, it remains cumbersome to collect documents covering the same event and rebuild their original linkage.

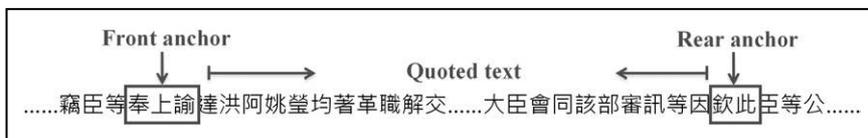
One of the important relations among the Imperial Court documents is the citation links among the Imperial Edicts and the Memorials. A Memorial often quotes earlier Imperial Edicts as the directives for the action being reported, while an Imperial Edict may cite earlier Memorials as the reason for issuing the decree. In an important historical event such as the Taiping Rebellion, there may be hundreds of Imperial Edicts and Memorials that form a complex web of successive citations. We call such a graph an *IE-M diagram*. Figure 1 is an illustration. Such a diagram often reveals some of the details of how the Qing court dealt with a specific event.



**Figure 1.** An abstracted IE-M Diagram.

## 2.2 Our Method

To check if a document has cited previous documents is similar to plagiarism detection (or copy detection), which detects whether a document contains portions copied from other materials (without acknowledging the source)<sup>9</sup>. The usual approach is to generate ‘fingerprints’ of a document by randomly selecting substrings from the text and then exhaustively comparing them with the fingerprints of all the other documents stored in the database. Such comparisons require significant computational resources. Thus the major research emphasis on this problem has been to design strategies that require fewer fingerprints in order to speedup the computation<sup>10</sup>.



**Figure 2.** In a memorial, the sentences in between the syntactic anchors are text quoted verbatim from the source edict.

An important observation of the Imperial Court documents is that there are often specific phrases around quotations that the authors used to prompt their readers about the presence of quotations. We call such phrases *syntactic anchors*. For example, in the case of a Memorial citing an Imperial Edict, the quotation in the former often starts with a *front anchor* ‘奉上諭’ (*adhering to the Imperial Edict*) and ends with a *rear anchor* ‘欽此’ (*By the Emperor Himself. That is all.*). The text in between is quoted verbatim from the Edict (although may not be in its entirety) (see Figure 2). Similarly, in the case of an Imperial Edict citing a Memorial, there are syntactic anchors indicating the presence of quotations. However, the anchors used in Imperial Edicts have a number of varieties and the quotations may also be done in a more casual manner (rather than being quoted verbatim). We have collected pairs of syntactic anchors for locating the start and end positions of quotations in both cases. We also remark that not all quotations have both front and rear anchors. However, since our algorithms do not require the entire quoted text, the absence of one anchor does not hinder their effectiveness.

After identifying the anchors and the quoted text, our method extracts a segment of the latter (called a *signature*) and applies a text-matching algorithm to see if it appears in any document in the database. Let  $n$  be the length of the signature. In the case of an Edict citing a Memorial,  $n$  is chosen to be 15 words, and for a Memorial citing an Edict,  $n$  is 20.

We use two different text-matching algorithms. In the case of a Memorial citing an Imperial Edict, since imprecision may occur due to the use of synonyms or transcription error, we use a string matching algorithm that tries to match documents which contain a substring  $s'$  such that the edit distance between the signature and  $s'$  is no larger than  $r$ , a predefined integer indicating the number of different words to be tolerated<sup>11</sup>. A more detailed description is given in Figure 3.

The algorithm for the case of an Edict quoting a Memorial is more complicated, since the quoted text might be a combination of segments selected from the cited Memorial and the orders of segments might not have been maintained. In our algorithm, we treat the signature as a sequence of  $n-1$  bi-grams (2-word sequences). Given a document  $d$  in the document set, we compute the percentage of bi-grams appeared in  $d$ . Any document that contains bi-

Given a Memorial  $d$  with signature  $s = \langle w_1, w_2, \dots, w_n \rangle$ , a string of words, let  $d'$  be a document.

Then  $d'$  is a candidate if  $d'$  contains a substring  $s'$  such that the edit distance of  $s$  and  $s' \leq r$ .

In our experiment, we set  $n = 15$  and  $r = 1$ .

**Figure 3.** The text-matching algorithm for Memorials citing Imperial Edicts.

Given an Imperial Edict  $d$  with signature  $s = \langle w_1, w_2, \dots, w_n \rangle$ , let  $d'$  be a document.

Define  $n-1$  bi-grams  $b_1 = \langle w_1, w_2 \rangle, \dots, b_i = \langle w_i, w_{i+1} \rangle, \dots, b_{n-1} = \langle w_{n-1}, w_n \rangle$  and two functions

$$presence(b_i, d') = 1 \text{ if } b_i \text{ occurs in } d'$$

$$presence(b_i, d') = 0 \text{ if } b_i \text{ does not occur in } d'$$

and

$$score(s, d') = \frac{\sum_{i=1}^{n-1} presence(b_i, d')}{n-1}$$

Then  $d'$  is a candidate if  $score(s, d') \geq t$  where  $t$  is a predefined threshold.

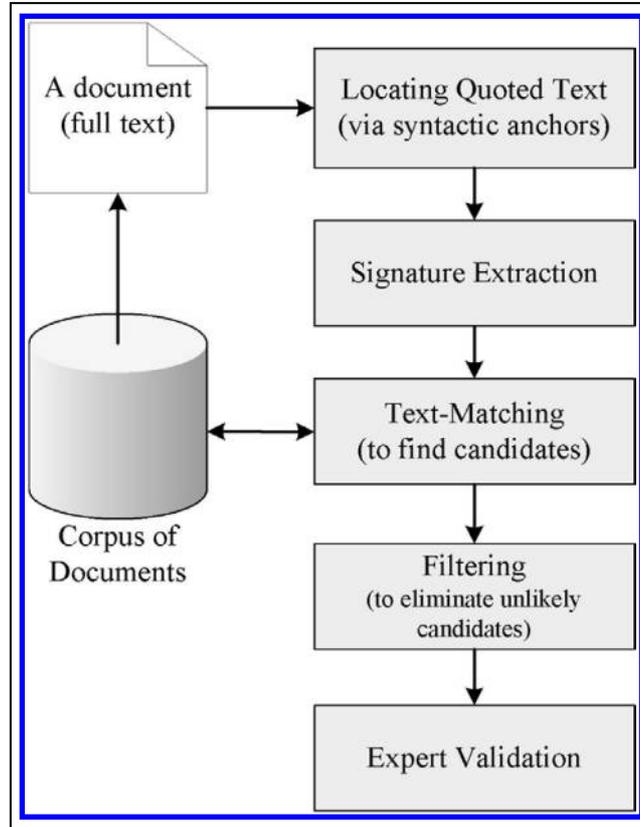
In our experiment, we set  $n = 20$  and  $t = 0.6$ .

**Figure 4.** The text-matching algorithm for Imperial Edicts citing Memorials.

grams over a threshold is considered a candidate. A more detailed description is shown in Figure 4.

When a document contains a substring that matches the signature, it is labelled as a *candidate*. There may be several candidates returned from the algorithm for the same signature. A filtering process is then applied to weed out those that are incorrectly returned. These include documents that may have quoted the same source or were completed at a later date.

Our method produces a list of  $\langle d, candidate \rangle$  pairs. We then present the findings to historians for manual validation to ensure accuracy. The overall process of our approach is shown in Figure 5.



**Figure 5.** The overall process of our method for the discovery of citation relations.

### 2.3 Experimental Results

We applied our method to the 37,831 Taiwan-related Imperial Court documents in THDL, and discovered 8,993 potential citation pairs, among which 6,490 pairs are possible Memorials citing Edicts, and 2,503 pairs are possible Edicts citing Memorials. We then asked historians to validate their correctness. Of the 6,490 pairs of potential Memorials citing Edicts, 4,883 are found to be correct, and of the 2,503 pairs of Edicts citing Memorials, 1,918 are correct. That puts the precision rates at 75.2 per cent and 76.6 per cent, respectively (see Table 1). What is also significant is that about 80–90 per cent of these citation pairs are from different sources, which would have been difficult to find without a collection such as THDL and an automated method such as presented here.

**Table 1.** The result of our method for the discovery of citation relations.

	Memorials citing Edicts	Edicts citing Memorials
Citation relations found ( $\langle d, candidate \rangle$ pairs)	6,490 pairs	2,503 pairs
Correct pairs	4,883 pairs	1,918 pairs
Precision	75.2%	76.6%
Pairs from different sources	4,614 (94.5%)	1,580 (82.4%)

**Table 2.** A summary of the sizes of the IE-M diagrams produced by our method.

Diagram size	Number of diagrams
153	1
126	1
109	1
107	1
95	1
85	1
83	1
62	1
40–49	5
30–39	3
20–29	15
10–19	70
5–9	283
4	164
3	249
2	304
<b>Total</b>	<b>1,101</b>

#### 2.4 The IE-M Diagrams

By taking the transitive closure (connecting successive citations) of the citation pairs, we produced 1,101 IE-M diagrams, among which 101 involve more than 10 documents (see Table 2). To help historians observe these diagrams, we used Graphviz<sup>12</sup>, a visualisation software, to represent them. Figure 6 is an example of such a diagram, in which the clear blocks represent an Imperial Edicts and the shadow ones represent Memorials. The arrows between the blocks indicate citations. Figure 6 shows correspondences between Emperor Qianlung (乾隆) and Kuilun (魁倫), the governor of the Fujian (福建) Province, discussing about how to deal with anticipated grain shortage in Fujian due to an earlier draught in Taiwan, the major source of imported grain. The block at the top is a Memorial

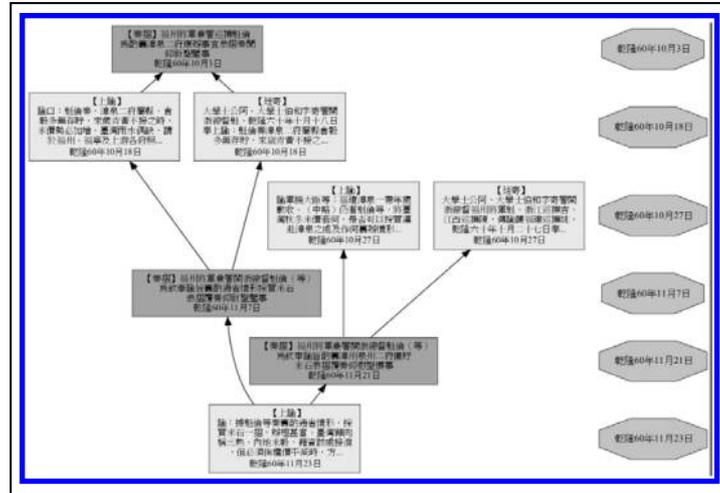


Figure 6. An example of IE-M diagram.

from Kuilun reporting this issue and his plan to buy grains from neighbourhood prefectures. The following two Imperial Edicts are opinions from Qianlung and instructions to further investigate the grain prices in the neighbourhood areas. The Edicts were then followed by Kuilun’s report on the grain prices and his revised plan. By tracing citation links in this diagram, we can vividly see the details of how a policy evolved through the discussions among the emperor and his officials and then carried out.

We also remark that all of the larger IE-M diagrams (size over 40) are about major events in Taiwanese history (see Table 3), several of which corresponded to different stages of the Lin Shuangwen (林爽文) Rebellion, the largest civil unrest in Taiwan during the Qing reign. The diagrams (such as the one shown in Figure 7) vividly reflected how the rebellion, first dismissing as a minor local disturbance, developed into an island-wide revolt. (The rebels even overran a prefectural seat and had another under siege for more than six months). It also showed how the local officials, failed at suppressing the revolt, pointed fingers at each other or fabricated victories. Emperor Qianlung finally realised the severity of the situation and sent Fukangan (福康安), one of his most able generals, to put down the rebellion. (Qianlung himself considered the pacification of Lin Shuangwen Rebellion one of his ‘Top Ten Military Achievements’.) As a prologue, one diagram (of size 95) showed a governing policy change that lasted more than 100 years in response to the aftermath of the Lin Shuangwen Rebellion.

**Table 3.** The main content of the IE-M diagrams sized over 40.

<b>Diagram size</b>	<b>Main Content</b>
153	Edicts and Memorials about the Lin Shuangwen Rebellion (1786–88). This diagram contains mainly correspondences between Emperor Qianlong and Chang Qing (常青) and other officials from the 6 <sup>th</sup> month of the 52 <sup>th</sup> year to the 2 <sup>nd</sup> month of the 53 <sup>th</sup> year of Qianlong, concerning the siege of the city of Zhuluo (諸羅). It also contains Edicts urging Fukangan (福康安), the newly appointed chief military commander, to go to Taiwan as soon as possible.
126	This diagram involves the Edicts and Memorials from the 6 <sup>th</sup> month to the 12 <sup>th</sup> month of the 52 <sup>th</sup> year of Qianlong during the Lin Shuangwen Rebellion, concerning mainly with the requests from Chang Qing to move troops from nearby provinces to support Taiwan.
109	Edicts and Memorials from the 1 <sup>st</sup> month to the 6 <sup>th</sup> month of the 52 <sup>th</sup> year of Qianlong, the early stage of the Lin Shuangwen Rebellion. This diagram includes the very first memorial from Chang Qing reporting the fall of Changhua (彰化) and Zhuluo, and details about the dispatching of troops and the situation on the battlefields.
107	Edicts and Memorials from the 4 <sup>th</sup> month of the 52 <sup>th</sup> year to the 5 <sup>th</sup> month of the 53 <sup>th</sup> year of Qianlong during the final stage of the Lin Shuangwen Rebellion. This diagram contains Qianlong's enquiries about Chai Daji (柴大紀), a military officer originally hailed as a war hero then condemned as corrupted and a coward who falsely claimed victories.
95	When dealing with the aftermath of the Lin Shuangwen Rebellion, Qianlong declared that all new magistrates of Taiwan Dao (台灣道) be given the title of Anchashe (按察使) so that they can submit Memorials to the emperor directly. This Imperial Edict was cited in the first Memorial of every new magistrate of Taiwan Dao until 1867 (1787–1867)
85	Edicts and Memorials about the raids of Cai Qian (蔡牽), a Chinese pirate, and his subsequent capture (1806–09)
83	Edicts and Memorials about the Mudan She (牡丹社) Incident, an expedition launched by the Japanese in retaliation for the killing of 54 Ryukyuan sailors by Paiwan aborigines near the southwestern tip of Taiwan (1874–75)
62	Edicts and Memorials regarding the invasion of Keelung and Tasui and the subsequent blockade of northern Taiwan by the French during the Sino-French War (1884–85)

Table 3. Continued.

Diagram size	Main Content
47	Sent Shen Baozhen (沈葆楨) to deal with the Japanese who went to Su-ao after the Mudan She Incident (1874–1875)
45	Reports from Chang Qing during the early stage of the Lin Shuangwen Rebellion (1786–87)
45	Edicts and Memorials concerning the incident of Nerbudda and Ann, two British vessels sank off the coast of Taiwan during the First Opium War (1839–42)
43	Pacification of the indigenous people by Liu Mingchuan (劉銘傳) (1885–90)
42	Edicts and Memorials about the establishment of the Taiwan Province (1885–88)

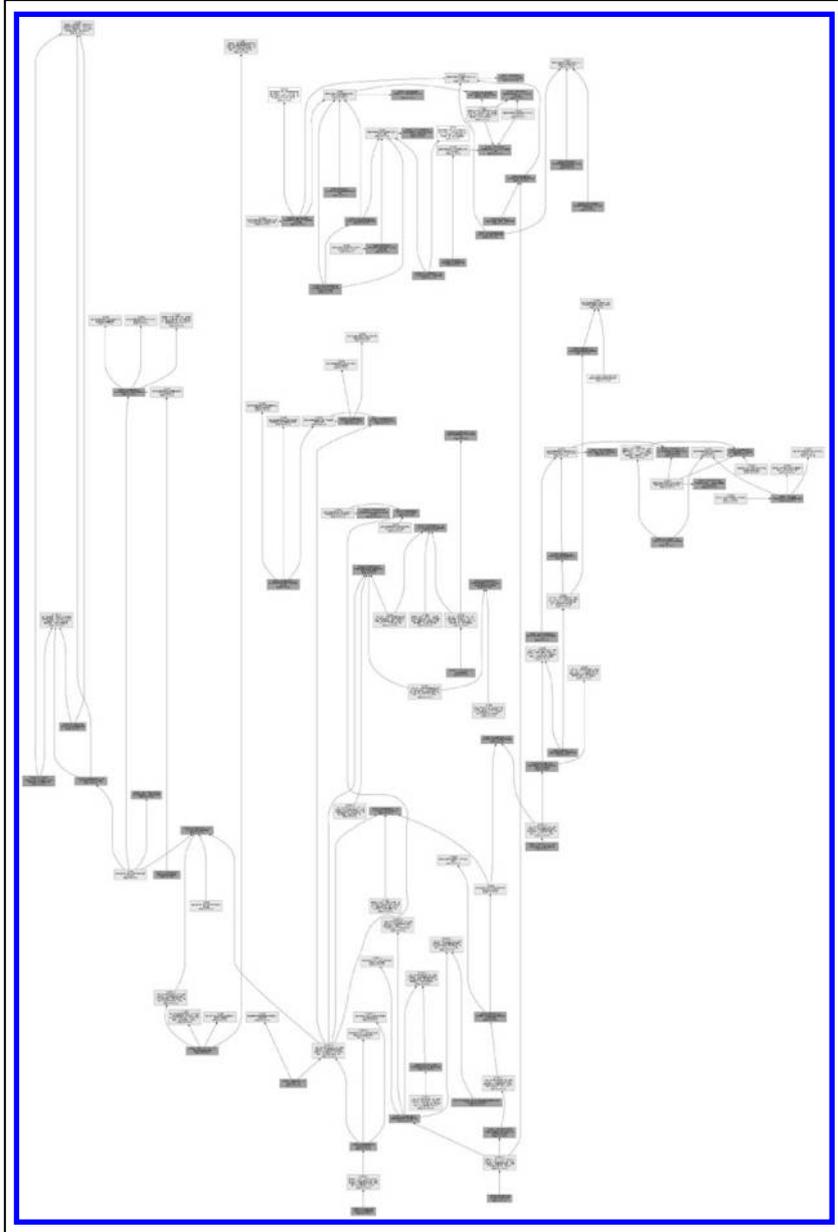
## TEMPLATE DISCOVERY

## 3.1 Motivation

Since the Imperial Court documents of THDL were selected from 235 different original sources, we were also interested to see if there were duplicates that were extracted from different sources. The similarity measure we used is based on the notion of finding the *longest common subsequences (LCS)* to reflect the amount of identical parts shared by two documents<sup>13</sup>. Our similarity measure is defined as:

$$\text{similarity}(d_1, d_2) = \frac{2 \times \text{LCS}(d_1, d_2)}{\text{wordcount}(d_1) + \text{wordcount}(d_2)} \quad (1)$$

By computing similarity between each pair of documents, we found 38,357 pairs that share over 80 per cent of the same texts. We then grouped highly similar documents together and obtained 3,973 groups. Examining the contents of each group revealed something quite interesting. The causes of similarity in the Imperial Court documents can be divided into three types: *duplicates*, *transcriptions* and *templates*. Duplicates are what we had anticipated when starting this investigation, that the same documents being included in different sources. Transcriptions are different versions of the same document that had been transcribed and kept at difference offices during the production process. A typical example is a Memorial that went through different offices and was transcribed (sometimes also modified) at each office. The most surprising type, however, is the templates, which are documents written following a similar format. Table 4 lists the sizes of the 3,973 groups of similar documents



**Figure 7.** The largest IE-M diagram, involving 153 Imperial Edicts and Memorials, is about the Lin Shuangwen Rebellion, the largest civil unrest in Taiwan during the Qing reign.

**Table 4.** The groups of similar documents discovered in the Imperial Court documents of THDL.

<b>Documents Involved</b>	<b>Number of Groups</b>	<b>The Causes for Similarity</b>
221	1	Templates
212	1	Templates
90	1	Templates
56	1	Templates
54	2	Templates
10–15	13	Templates (mostly), Duplicates, Transcriptions
5–9	97	Duplicates, Transcriptions
3–4	741	Duplicates, Transcriptions
2	3,116	Duplicates, Transcriptions
<b>Total</b>	<b>3,973</b>	

discovered in the Imperial Court documents, along with their causes of similarity. We found that the large (sized over 50) and medium sized groups (10 to 15) are all templates, while the small groups (sized below 9) are duplicates or transcriptions.

### 3.2 *Templates*

Table 5 shows the list of top 7 templates we discovered and brief descriptions of their contents. The largest template contains 221 monthly reports from the local officials to the emperors listing the prices of grains in the 11 prefectures of the Fujian Province. Such a report first gave the name of the reporting official, followed by the month being reported, then listed the prices of rice in three tiers of quality in each of the 11 prefectures in that month. The major differences among the 221 documents are only at the name of the reporting official, the month being reported, and the prices (see Figure 8).

The second largest template, 212 in total, are memorials reporting the numbers of Jiansheng (監生, students in the Imperial Academy) titles purchased each month in Fujian. All reports started with a brief summary of the sum of silver that had been accrued prior to the month reporting, how the money had been used, and the balance. The reports then gave the numbers of Jiansheng titles purchased in the reported month along with the sum of silver received. Figure 9 shows such a report.

Two other smaller groups, sized 54 and 15, are also reports of the numbers of Jiansheng purchased but follow somewhat different formats. Figure 10 shows an example from the template sized 15, whose format is almost the same with the report in Figure 9 except for two additional reasons added to explain how the

**Table 5.** The large templates discovered in the Imperial Court documents of THDL.

<b>Group Size</b>	<b>Contents</b>
221	Reports of the prices of grains in the major cities of the Fujian Province
212	Reports of the number of Jiansheng (監生) titles purchased in the Fujian Province
90	Reports listing the customs tax received in the very first two ports opened for foreign trade, Huwei (滬尾) and Dagou (打狗), in Taiwan.
56	Listings of candidates of buzhengshi (布政使, the Provincial Administration Commissioner)
54	Reports of the numbers of Jiansheng (監生) titles purchased in the Fujian Province (a slightly different template from the above)
54	Reports listing the customs tax received in the very first two ports opened for foreign trade, Huwei (滬尾) and Dagou (打狗), in Taiwan.
15	Reports of the numbers of Jiansheng (監生) titles purchased in the Fujian Province (slightly different from the above)

money was used. We also found that the changes of templates were sequential. The template with the largest number of reports (shown in Figure 9) was used from the 13<sup>th</sup> to the 30<sup>th</sup> year of the Daoguang (道光) reign (1833–1850). Then the format changed to the template sized 15 (shown in Figure 10) when the Xianfeng (咸豐) reign began in 1851. Then in 1856, during the 6<sup>th</sup> year of Xianfeng, the format changed again and became the template sized 54. The two templates, sized 90 and 54, in Table 5 are reports listing the custom taxes received in Huwei (滬尾) and Dagou (打狗), the first two ports in Taiwan opened for foreign trade. They also demonstrated the change of report format over time: the first group was from the 8<sup>th</sup> year of Tongzhi (同治) to the 14<sup>th</sup> year of Guangxu (光緒), 1869–1888, and the second from the 14<sup>th</sup> year of Guangxu to the 20<sup>th</sup>, 1888–1894. These two sets of reports covered almost the entire period from the opening of the treaty ports in Taiwan (1861) to Taiwan’s secession to Japan (1895).

The official	The month	
福建巡撫田黃檢謹奏	謹將福建省乾隆四十四年正月分	米糧時價
敬繕清單	恭呈御覽	
福州府屬糧價比上月稍增 (價中) :	上米每倉石價銀一兩六錢至二兩一錢三分 ; 中米每倉石價銀一兩五錢三分至一兩九錢六分 ; 下米每倉石價銀一兩四錢五分至一兩八錢	← The prices of the best, medium, low quality of rice in the Fuzhou prefecture
臺灣府屬糧價與上月相同 (價中) :	上米每倉石價銀一兩三錢至一兩五錢五分 ; 中米每倉石價銀一兩二錢至一兩四錢五分 ; 下米每倉石價銀一兩一錢六分至一兩三錢五分	← The prices in the Taiwan prefecture
.....		
	暫護福建巡撫奴才伊爾布	謹奏
	謹將福建省乾隆五十三年十月分各屬米糧價值	敬繕清單
	恭呈御覽	
福州府屬 (價中) 與上月相同 :	上米每倉石銀一兩四錢五分至二兩三錢 ; 中米每倉石銀一兩三錢八分至二兩二錢 ; 下米每倉石銀一兩三錢至二兩八分	←
臺灣府屬 (價中) 比上月稍減 :	上米每倉石銀二兩至三兩二錢 ; 中米每倉石銀一兩七錢至二兩六錢五分 ; 下米每倉石銀一兩四錢至二兩二錢	←
.....		

**Figure 8.** Two memorials listing the prices of grains of Fujian. The major differences of the two are only at the names of reporting officials, the months being reported, and the prices.

### 3.3 Extracting tabular data from templates

In a template where a group of documents follow the same format, the differences among the documents usually contain the most significant information. For example, in the group containing 221 monthly reports on the grain price in Fujian, the differences in the reports are only the dates and the prices. It is natural then to extract the information and turn them into tabular data. We wrote a program that produced the extracted numerical data, a portion of which is presented in Figure 11.

Figure 12 gave a better visualisation of the fluctuation of rice prices between the 3<sup>rd</sup> month of the 48<sup>th</sup> year and the 10<sup>th</sup> month of the 55<sup>th</sup> year of Qianlung (1783–1790). Note the conspicuous absence of reports during and after the Lin Shuangwen Rebellion (between the 5<sup>th</sup> month of the 49<sup>th</sup> year and the 7<sup>th</sup> month of the 53<sup>rd</sup> year). Also note that when the report resumed, 6 months after the rebellion ended, the price of rice was twice of that before the rebellion. However, the price dropped significantly immediately after the first harvest (although did

The sum of the money that have been received from the purchasing during the period	The period (since the time the Jiansheng purchase began up to the prior month)	
<p>再，福建省收捐監生銀數，自嘉慶五年三月起，至道光三十年十一月止，共收捐銀三百七十八萬六千一百九十兩，除節次解部並歸補本省司庫封存同臺灣軍需防夷經費動用暨撥臺灣道庫以及撥解東河並江南河庫工需又續撥澳省銅本外，尚存銀三萬三千五百八十八兩；又自道光二十二年二月起，至三十年十一月止，共收捐武監生銀三萬五千六百四十兩，內除撥澳解赴東河並江南河庫工需又續撥澳省銅本共銀三萬二千五百一十八兩外，尚存銀三千一百二十二兩，經臣附片奏明在案。</p>		Where the money has gone
<p>The present month</p>		The balance prior to the present month
<p>茲據司詳，道光三十年十二月分收捐文監生三十三名，計收銀三千五百六十四兩，連前共存銀三萬七千一百五十二兩，俟收有成數，即行撥補二十八年分臺餉失水銀四萬九千八百八十一兩六錢六分三釐八毫釐除，另再詳請奏報；又十二月分收捐武監生三名，計收銀三百二十四兩，連前共存銀三千四百四十六兩，等由。前來。臣覆查無異，理合附片具奏，伏乞聖鑒。謹奏。</p>		The amount of Jiansheng titles (Literat) purchased and the amount of money received
<p>The amount of Jiansheng titles (Military) purchased and the amount of money received</p>		

Figure 9. A memorial from the largest template (sized 212) about Jiansheng.

<p>再，福建省收捐監生銀數，自嘉慶五年三月起至咸豐元年八月止，共收捐銀三百八十三萬九千七百五十八兩，內除節次解部，並歸補本省司庫封存，同臺灣軍需防夷經費動用，暨撥臺灣道庫以及撥解東河並江南河庫工需，又續撥澳省銅本，暨奉准作正開銷，補發道光二十八年分臺餉失水，並准部指撥廣西省軍需外，尚存銀三千七百七十四兩三錢三分六厘二毫，又自道光二十二年二月起至咸豐元年八月止，共收捐武監生銀三萬八千四百四十八兩，內除撥澳解赴東河並江南河庫工需，又續撥澳省銅本，共銀三萬二千五百一十八兩外，尚存銀五千九百三十兩，業經臣附片奏明在案。</p>	Two more reasons were added to explain where the money has gone
<p>茲據司詳，咸豐元年閏八月分，收捐文監生四十二名，計收銀四千五百三十六兩，連前共存銀八千三百一十兩三錢三分六厘二毫，內除續撥廣西省軍需銀八千兩外，實尚存銀三百一十兩三錢三分六厘二毫，仍俟收足三萬兩，歸補本省封儲。又閏八月分，收捐武監生一十七名，計收銀一千八百三十六兩，連前共存銀七千七百六十六兩，內除續撥廣西省軍需銀二千兩外，實尚存銀五千七百六十六兩。等由。前來。臣履核無異，理合附片具奏，伏乞聖鑒。謹奏。</p>	

Figure 10. An example document from the template sized 15 about Jiansheng. Its format is almost the same with the one shown in Figure 9 but differs on explaining how the money was used.

Year	1783	1784	1785	1786	1787	1788	1789	1790
泉州府(中米)	2.15	2.075	1.975	2	1.925	1.825	1.85	1.75
漳州府(中米)	2.15	2.075	1.975	2.025	1.95	1.85	1.8	1.7
興化府(中米)	2.15	2.075	1.975	2.075	1.975	1.875	1.84	1.74
龍巖州(中米)	2.075	1.975	1.875	1.975	1.875	1.825	1.815	1.74
臺灣府(中米)	2.075	1.975	1.875	1.975	1.875	1.825	1.815	1.74
福州府(中米)	2.075	1.975	1.875	1.965	1.865	1.815	1.79	1.705
汀州府(中米)	2.15	2.075	2	1.965	1.865	1.815	1.79	1.705
福寧府(中米)	2.36	2.285	2.185	2.285	2.185	2.135	2.115	2.025
延平府(中米)	2.45	2.375	2.275	2.375	2.275	2.225	2.205	2.115
建寧府(中米)	2.5	2.425	2.325	2.425	2.325	2.275	2.255	2.165
邵武府(中米)	2.36	2.285	2.185	2.285	2.185	2.135	2.115	2.025
泉州府(中米)	2.15	2.075	1.975	2	1.925	1.825	1.85	1.75
漳州府(中米)	2.15	2.075	1.975	2.025	1.95	1.85	1.8	1.7
興化府(中米)	2.15	2.075	1.975	2.075	1.975	1.875	1.84	1.74
龍巖州(中米)	2.075	1.975	1.875	1.975	1.875	1.825	1.815	1.74
臺灣府(中米)	2.075	1.975	1.875	1.975	1.875	1.825	1.815	1.74
福州府(中米)	2.075	1.975	1.875	1.965	1.865	1.815	1.79	1.705
汀州府(中米)	2.15	2.075	2	1.965	1.865	1.815	1.79	1.705
福寧府(中米)	2.36	2.285	2.185	2.285	2.185	2.135	2.115	2.025
延平府(中米)	2.45	2.375	2.275	2.375	2.275	2.225	2.205	2.115
建寧府(中米)	2.5	2.425	2.325	2.425	2.325	2.275	2.255	2.165
邵武府(中米)	2.36	2.285	2.185	2.285	2.185	2.135	2.115	2.025
泉州府(中米)	2.15	2.075	1.975	2	1.925	1.825	1.85	1.75
漳州府(中米)	2.15	2.075	1.975	2.025	1.95	1.85	1.8	1.7
興化府(中米)	2.15	2.075	1.975	2.075	1.975	1.875	1.84	1.74
龍巖州(中米)	2.075	1.975	1.875	1.975	1.875	1.825	1.815	1.74
臺灣府(中米)	2.075	1.975	1.875	1.975	1.875	1.825	1.815	1.74
福州府(中米)	2.075	1.975	1.875	1.965	1.865	1.815	1.79	1.705
汀州府(中米)	2.15	2.075	2	1.965	1.865	1.815	1.79	1.705
福寧府(中米)	2.36	2.285	2.185	2.285	2.185	2.135	2.115	2.025
延平府(中米)	2.45	2.375	2.275	2.375	2.275	2.225	2.205	2.115
建寧府(中米)	2.5	2.425	2.325	2.425	2.325	2.275	2.255	2.165
邵武府(中米)	2.36	2.285	2.185	2.285	2.185	2.135	2.115	2.025

Figure 11. A portion of the table obtained by extracting the crucial differences from the largest template, containing 221 monthly reports on the prices of grains in Fujian.

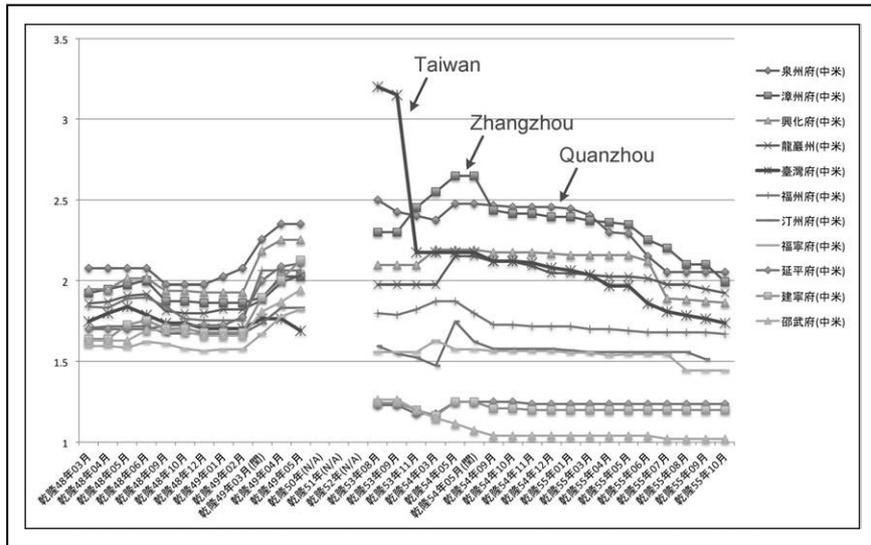


Figure 12. The line chart visualisation for the data in Figure 11 (1783–1790). Each line represents the rice prices in a prefecture in Fujian.

not go back to the pre-revolt level for another two years). It is also worth noting that Zhangzhou (漳州) and Quanzhou (泉州), two Fujian prefectures that had always relied heavily on rice imported from Taiwan<sup>14</sup>, actually saw the prices of their rice rise after the Taiwan rice harvest. This might be because these two prefectures received rice shipments from other regions during the war to keep their prices stable, but the shipment stopped after the Taiwan revolt was

Month	Original text for Literary Jiansheng	Original text for Military Jiansheng	No. of Literary Jiansheng	Subtotal	No. of Military Jiansheng	Subtotal
道光21年11月	武監生一十九名, 計收銀二千五百一十六兩	武監生五十二名, 計收銀五千六百一十六兩	19	2052	2	216
道光21年12月	文監生五十九名, 計收銀五千四百九十二兩	武監生五十四名, 計收銀五千八百五十二兩	59	5400	9	972
道光21年1月	文監生七十四名, 計收銀七千九百九十二兩	武監生七十二名, 計收銀七千七百七十二兩	74	7987	12	1296
道光21年2月	文監生五十二名, 計收銀五千六百一十六兩	武監生八十四名, 計收銀九千零七十二兩	52	5416	9	972
道光21年3月	文監生五十六名, 計收銀六千四百八十八兩	武監生九十三名, 計收銀一萬四千四百兩	56	6088	14	1512
道光21年4月	文監生六十一名, 計收銀六千五百八十八兩	武監生一百七名, 計收銀一萬一千五百五十六兩	61	6589	2	216
道光21年5月	文監生一百一十三名, 計收銀一萬二千二百四兩	武監生一百九名, 計收銀一萬一千七百七十二兩	113	12204	9	972
道光21年6月	文監生五十二名, 計收銀五千六百一十六兩	武監生一百二十名, 計收銀一萬四千四百兩	52	5416	8	864
道光21年7月	文監生五十八名, 計收銀六千二百六十四兩	武監生二名, 計收銀二百一十六兩	58	6084	2	216
道光21年8月	文監生四十八名, 計收銀五千一百八十四兩	武監生三名, 計收銀三百二十四兩	48	5384	3	324
道光21年9月	文監生九名, 計收銀九百七十二兩	武監生一名, 計收銀一百八兩	9	972	1	108
道光21年10月	文監生四十四名, 計收銀五千四百九十二兩	武監生三名, 計收銀三百二十四兩	44	4752	3	324

Figure 13. A portion of the table obtained by extracting the crucial differences from the three templates about Jiansheng in Fujian, totaled 278 memorials.

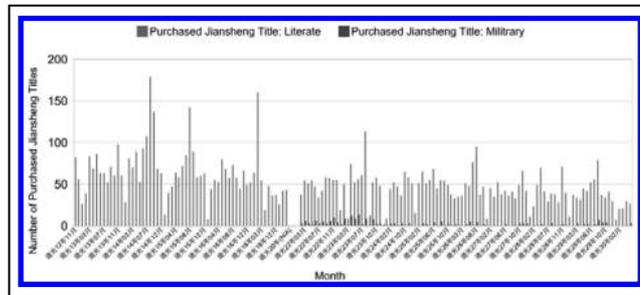


Figure 14. The purchase of literate and military Jiansheng titles in the Fujian province (1833–1850).

over on the assumption that the regular flow of rice from Taiwan would resume. However, rice from Taiwan did not arrive as anticipated, which caused the prices to rise again.

In the case of Jiansheng titles purchased Fujian (totalling 278 documents), we have also extracted the numbers purchased each month (both literate and military Jiansheng) and the total amount of silver reported to the emperors. A portion of the results is shown in Figure 13. Dividing the total sum of each year by the number of titles purchased revealed that the official price for such a title is 108 liang (兩) silver, which matches what other studies showed<sup>15</sup>.

The data can also be presented on a bar chart, as Figure 14 shows those between the 13<sup>th</sup> year and the 30<sup>th</sup> year of Daoguang (1833–1850). Note that the purchase of military Jiansheng (marked in shadow) only appeared after 1842, the year the First Opium War (1839–1842) ended. Another phenomenon is that there seems to be a seasonal pattern in the number of Jiansheng purchased, with peaks in summer (7<sup>th</sup> month of the year) and valleys in winter (1<sup>st</sup> month).

CONCLUSION

With digital resources become more and more common place, it becomes imperative for researchers to use information technology to effectively explore the digital archives. Although search engines and cross-platform search tools may let the user find a large amount of material, the overwhelming quantity sometimes can become more of a nuisance than help. A major reason is because search systems often treat the returned documents as unrelated. Thus the user must examine all of the retrieved documents if he wants to know what have been found. THDL<sup>16</sup>, the Taiwan History Digital Library, is built on the hypothesis that historical documents are often interrelated, and that the system should help the user explore the potential relationships. Thus, we have developed a methodology that treats a query return as a sub-collection and have designed an environment in which the user can observe the *collective meaning* of the returned sub-collection<sup>17</sup>. We have also developed IT methods that discover important relationships that are hidden among the documents. Two of them, the *citation relation* and the *templates*, among 37,831 court documents of Qing, are presented here<sup>18</sup>. The transitivity closures of the citation relations are further captured in a concept of IE-M diagram, which vividly reveals how a historical event evolved through the correspondences between a Qing emperor and his officials. The templates show how some reports followed certain formats. In some cases, the numbers involved (such as grain prices at Fujian) can be extracted and tabulated.

Our methods demonstrate how to use information technology to explore and identify important relations among historical documents that would be hard to find otherwise. The main purposes of our tools are to discover new contexts and to provide an environment for users to explore them.

END NOTES

<sup>1</sup> S. P. Chen, J. Hsiang, H. C. Tu, and M. C. Wu, 'On building a full-text digital library of historical documents', in *Lecture Notes in Computer Science: Proceedings of the 10th international conference on Asian digital libraries: looking back 10 years and forging new frontiers, Hanoi, 2007*, 4822 (2007), 49–60.

<sup>2</sup> M. C. Wu, K. I. Ang, W. L. Lee, and H. Y. Lin, *A brief introduction to the integrated collections of Taiwan-related historical records* (Taiwan, 2005).

<sup>3</sup> Details about post-query classification and co-occurrence analysis can be found in S. P. Chen, J. Hsiang, H. C. Tu, and M. C. Wu, 'On building a full-text digital library of historical documents', 49–60; and S. P. Chen, J. Hsiang, and H. C. Tu, 'On building a full-text Digital Library of Land Deeds of Taiwan', *Conference Abstracts of the 2009 Digital Humanities Conference, College Park, Maryland*. (Maryland, 2009), 85–90.

<sup>4</sup> Chien Mu, *Chung-kuo li tai cheng chih te shih* (Hong Kong, 1952).

<sup>5</sup> J. F. Chuang, *Qing dai shi liao lun shu* (Taipei, 1979).

- <sup>6</sup> ‘Database of Ch’ing palace memorials and archives of the grand council’, *National Palace Museum* [World Wide Web homepage]. < URL: <http://npmhost.npm.gov.tw/tts/npmmeta/dblist.htm> > [12 Mar 2011].
- <sup>7</sup> Institute of History and Philology, Academia Sinica, *The archives of the grand secretariat in Academia Sinica* [World Wide Web homepage]. < URL: <http://archive.ihp.sinica.edu.tw/mctkm2/index.html> > [12 Mar 2011].
- <sup>8</sup> The First Historical Archives of China (Zhongguo diyi lishi dang’an guan), *Zhu pi zou zhe in the Guangxu’s reign* (Beijing, 1995).
- <sup>9</sup> N. Shivakumar and H. Garcia-Molina, ‘SCAM: A copy detection mechanism for digital documents’, unpublished paper presented at the 2nd International Conference in Theory and Practice of Digital Libraries, Austin, Texas, 1995; A. Si, H. V. Leong, and R. W. H. Lau, ‘CHECK: a document plagiarism detection system’, *Proceedings of the 1997 ACM symposium on applied computing, San Jose*. (New York, 1997), 973–982; and S. Brin, J. Davis, and H. Garcia-Molina, ‘Copy detection mechanisms for digital documents’, *Proceedings of the 1995 ACM SIGMOD international conference on management of data, San Jose*. (New York, 1995), 398–409.
- <sup>10</sup> C. H. Timothy and Z. Justin, ‘Methods for identifying versioned and plagiarised documents’, *Journal of the American Society for Information Science and Technology*, 54:3 (2003), 203–215; and A. Chowdhury, O. Frieder, D. Rossman, and M. C. McCabe, ‘Collection statistics for fast duplicate document detection’, *ACM Transactions on Information Systems*, 20:2 (2002), 171–191.
- <sup>11</sup> For an explanation of edit distance, please see: < [http://en.wikipedia.org/wiki/Levenshtein\\_distance](http://en.wikipedia.org/wiki/Levenshtein_distance) >
- <sup>12</sup> < <http://www.graphviz.org/> >
- <sup>13</sup> T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, ‘Longest common subsequence’, *Introduction to algorithms* (Boston, 2003), 350–355.
- <sup>14</sup> M. E. Xie, *A study on the rice price on Taiwan in the Qing era (Qing dai Taiwan mi jia yan jiu)* Dao-Hsiang Publishing, Taipei, 2008.
- <sup>15</sup> Y. Wu, ‘A study of the purchase of an official rank in the Qing Dynasty’, *A Collection of Essays on the Ming and Qing Dynasties, Vol. 6* (Beijing, 2005), 4–27.
- <sup>16</sup> < <http://thdl.ntu.edu.tw> >
- <sup>17</sup> Details about this concept can be found in: S. P. Chen, J. Hsiang, H. C. Tu, and M. C. Wu, ‘On building a full-text digital library of historical documents’, 49–60; and S. P. Chen, J. Hsiang, and H. C. Tu, ‘On building a full-text digital library of land deeds of Taiwan’, 85–90.
- <sup>18</sup> For more tools, see < <http://thdl.ntu.edu.tw/tools/> >