



# 《引得市》的介紹與研究應用

陳信良

台北科技大學文化事業發展系兼任講師

## 摘要：

《引得市》創建於 2012 年 7 月，主要內容係古文字相關的字辭典、訓詁等索引資料庫。過去，我們利用這樣龐大的數據資料，進行了許多古文字演進的研究。希望藉由這樣的研究應用與介紹推廣，讓更多人了解使用，並且發展出更多可能的研究方式。

關鍵字：引得市 古文字 文獻 索引 數位化 大數據 應用 字典 辭典 文字編

## 一、前言

《引得市》<sup>1</sup>創建於 2012 年 7 月，主要內容係筆者與其他研究者共同製作的古文字相關字辭典索引資料。從紙本書籍數位化製作，再轉成資料庫提供網路檢索使用。因資料庫以書籍檢字內容為主，過去以來 INDEX 英文諧音為「引得」，INDEX 日本稱為「索引」(さくいん)。自取「引得」一詞，筆者增添「市」字，定名為《引得市》。或許也有「引德士」之諧音。「市」有眾多含意，引《老子》「美言可以市尊。」「換取」之意。研究者查書時，若能使用便利的電子索引檔、甚至資料庫，便利之餘，就能有更多研究的時間。<sup>2</sup>筆者感同身受，深知「時間」對研究者的重要性。因此《引得市》的存在，就是讓研究者能夠更有效率的作研究，減少研究中反復且不重要的過程。

近年的各項科學研究，多以「大數據」作為佐證或輔助說明。所謂「大數據」，維基百科的解釋是：大數據（Big data 或 Megadata），或稱巨量資料、海量資料、大資料，指的是所涉及的資料量規模巨大到無法透過人工，在合理時間內達到擷取、管理、處理，並整理成為人類所能解讀的形式的資訊。而累積諸多古文字相關文獻索引的《引得市》，就已經具備這樣的規模。

<sup>1</sup> 《引得市》網址：[www.mebag.com/index](http://www.mebag.com/index)

<sup>2</sup> 陳信良：〈文獻的索引數位化製作與應用研究-以「引得市」為例〉，刊：《書畫藝術學刊》第十三期，185-212 頁，臺北：國立臺灣藝術大學書畫藝術學系，2012 年版，第 187 頁。

---

《引得市》不僅提供文獻頁碼的索引，在登入最高權限的後端介面中，還能直接瀏覽文獻原始頁面，瞬間即可明瞭漢字字形結構、通假等各種資訊。2017年6月之後，結合了「開卷助理」<sup>3</sup>的應用，讓所有的使用者在檢索時，配合自己電腦中的文獻檔案，快速跳頁的瀏覽文獻。猶如最高權限的後端界面一樣功能。

這樣的考察方式在目前來說，應該是最快、最有效率的方式。經過大量的數據與文獻研究考察之後，能夠清楚理解「墨跡文字」與「非墨跡文字」的實質差異。可知其差異並非僅是「外形」與「材質」，關鍵應該是文字背後產生的方式，而差異也影響著討論「文字演進」時取用的分析討論。<sup>4</sup>《引得市》詳細內容數據與實際應用將於下段內容解說。

## 二、《引得市》的文獻種類與內容

《引得市》自2012年成立以來，平均每年增加約10種文獻。不定時更新文獻，累計製作245種文獻索引資料。<sup>5</sup>依文字的時間前後分成：「商周文字」「戰國文字」「秦漢文字」。以文字的時間不易區分的則分成：「古文字」「傳抄古文」「今文字」「說文」「康熙字典」「訓詁」「俗語詞」「辭典」「碑刻索引」「簡牘」「磚文」「璽印」「專書」「中日對照等項目。另外，「新編元代八斯巴字百家姓」「秦簡字形資料庫」「夢英篆書千字文」「古文字缺字資料庫」等內容，則不分類獨立呈現。以「古文字缺字資料庫」串連八大資料庫，支援資料庫中的電腦缺字顯示與部件的搜尋檢索。最新的文獻數量與字列請見下表一：

《引得市》提供查詢的，不僅只是字辭典、文字編的索引而已，也包含古文字通假這類的書籍，甚至還有一般的研究專書，例如：《當代中國簡帛學研究（1949-2009）》等。筆者製作這本書籍詳細索引的理由是：書中有各地出土的簡牘內容與數量明確整理與說明，雖然有些沒有註明簡牘資料的年代，也會附上原始出土的資料來源，讓研究者方便溯源考察。於欄位中輸入簡牘名稱查詢，或者直接列表顯示，都能包含筆者附加的時代分期，應該是接近完整的簡牘墨跡文字的列表了。

分類	商周文字	戰國文字	秦漢文字	古文字	傳抄古文	今文字
種類	14	29	56	20	14	24
字列	30213	55462	84690	62454	38573	95220

<sup>3</sup> 筆者blog拙文：〈建立「開卷助理」新的RMP檔step by step教學〉，網址：  
<https://ebag2007.blogspot.tw/2018/03/rmpstep-by-step.html>

<sup>4</sup> 陳信良：〈以墨跡文字為主的古文字演進研究方式〉，刊：《中國美術研究》，第18輯，83-90頁，南京：東南大學出版社，2016年版，第84頁。

<sup>5</sup> 最新數量請參考：INDEX引得市資料來源一覽表，括弧內為製作時間，網址：  
<https://ebag2007.blogspot.tw/2018/03/index.html>或「網路地圖」網址：  
[http://www.mebag.com/index/images/index\\_map.jpg](http://www.mebag.com/index/images/index_map.jpg)

分類	說文	康熙字典	訓詁	俗語詞	辭典	碑刻索引
種類	8	3	10	3	15	7
字列	9833	47043	64223	2850		

分類	簡牘	磚文	璽印	專書	中日對照
種類	7	1	1	30	4
字列					

表一 《引得市》文獻數量與字列

《引得市》決定文獻索引製作的優先次序，多數是筆者自己的研究考量，其他才是研究者、使用者所建議。因為有固定的一群朋友長期的使用與關注，共同編輯協作或相互交流補充文獻，持續累積才有現在的規模。

若沒有實際參與整個製作流程，一般人很難瞭解文獻數位化實際上所需要花費的時間。數位化文獻的時間概算<sup>6</sup>，以《楚文字編》為例，灰階 600dpi 的解析度，本文 910 頁、檢索表 34 頁、缺字 2158 字，計算的結果是：掃瞄時間 3.03 小時、檢字表製作 51 小時、缺字製作 17.98 小時。<sup>7</sup>一共需要約 72 小時。一天八小時，則需要約九個工作天才能完成。

由此可知，掃瞄所佔的時間並不是最多，主要是檢字表數位化的過程很花時間，無法在機器的效能上加快速度，必須由人工逐字檢查。我們使用的 OCR 文字辨識後再校正的方式，也比重新打字的方式快。檢字表的數位化，才能把字辭典發揮到極致。否則掃瞄後的圖檔，不容易搜尋，自然也不常使用，如此這些圖檔也只是佔據硬盤空間而已。當我們完成了字典檢字表的數位化，轉成資料庫來使用難度就不高了，在《引得市》就是最好的印證。<sup>8</sup>

多年前，筆者曾提出這樣的一種概念，學術著作即使不全文做成光碟，至少也把「索引頁」的文字檔，放在網路或者透過其他管道來分享使用。目前似乎也有學者付諸行動，劉志基先生說：「《中國出土簡帛文獻引得綜錄·包山楚簡卷》的附錄之一已經問世。與該紙質載體成果相配套，互聯網上亦將出現“《包山楚簡偏旁類纂》網路閱讀視窗。」<sup>9</sup>

<sup>6</sup> 筆者 blog 拙文：〈研究用書籍紙本數位化時間試算表 1.0〉，網址：  
<https://ebag2007.blogspot.tw/2012/08/10.html>

<sup>7</sup> 現在回頭看當初的時間預估，顯然也過度樂觀，缺字的製作的變數很多，時間遠比預估還更來的更久。

<sup>8</sup> 陳信良：〈文獻的索引數位化製作與應用研究-以「引得市」為例〉，刊：《書畫藝術學刊》第十三期，185-212 頁，臺北：國立臺灣藝術大學書畫藝術學系，2012 年版，第 200 頁。

<sup>9</sup> 劉志基：〈古文字偏旁數字平台與數字化環境下的古文字偏旁分析法〉，刊：《中國文字研究》，第 22 輯，1-9 頁，上海：上海書店出版社，2015 年版，第 2 頁。

《戰國古文字典：戰國文字聲系》是知名學者何琳儀先生的著作，為古文字研究相當重要的參考文獻。《引得市》成立之後，很想立即作索引的數位化，礙於原書是手寫字，OCR 文字辨識無法處理，重新打字又耗時費力，故一直沒有著手進行。2015 年時筆者有大量檢索使用的需求。下定決心動手處理。像這樣的文獻字頭索引數位化的進行，大約可分成三個階段：

- 1、依原書逐頁逐字增補，若為「缺字」則立即建立該字的「構字式」。(數天)
- 2、依需要的內容造字。(數天)
- 3、補充「筆畫檢字表」的頁碼及字頭的「筆畫」資訊。(數小時)

由此可知，缺字製作是文獻索引數位化過程中最耗時的。缺字造字得一字一字處理，沒有其他捷徑或直達車，為求字型風格一致，得由筆者獨立製作，無法假手他人。



The screenshot shows a Microsoft Excel-like spreadsheet interface titled '引得市資料來源'. The columns are labeled A through I. Column A is '資料名' (Title), B is '文獻名' (Name of the work), C is '文獻類型' (Type of work), D is '作者' (Author), E is '出版社' (Publisher), F is '出版時間' (Publication time), G is 'ISBN', H is '備註' (Notes), and I is '索引提供者' (Index provider). The data includes various books from the 1990s and 2000s, such as '甲骨文詞譜', '金文常用字典', and '新金文編', along with their respective details like authors like 朱岐祥, 陈初生, and 周法高, and publishers like 里仁書局 and 陝西人民出版社.

	A 資料名	B 文獻名	C 文獻類型	D 作者	E 出版社	F 出版時間	G ISBN	H 備註	I 索引提供者
1	甲骨文詞譜	甲骨文詞譜	書籍	朱岐祥	里仁書局	2013.12	978-9-866-17874-0		陳信良
2	甲骨文字典	甲骨文字典	書籍	徐中舒	四川辭書出版社	2006.09	7-80682-232-1	2版	趙撢的、陳信良
3	甲骨文字編	甲骨文字編	書籍	李宗焜	中華書局	2012.03	978-7-101-08134-3		小學堂、陳信良
4	甲骨文字詁林	甲骨文字詁林	書籍	于省吾	中華書局	1996.10	7-101-01430-5		小學堂、陳信良
5	甲骨文字詁林補編	甲骨文字詁林補編	書籍	何景成	中華書局	2017.10	978-7-101-12413-2		陳信良
6	金文常用字典	金文常用字典	書籍	陳初生	陝西人民出版社	2004.01	7-224-06658-3		彭偉明
7	金文詁林	金文詁林	書籍	周法高	香港中文大學	1974.06			小學堂、周旭
8	金文考釋匯纂(1990-1999)	金文考釋匯纂 ( 1990-1999 )	論文	蔣晴	吉林大學碩士論文	2017.05			於菟
9	金文考釋匯纂(2000-2015)	金文考釋匯纂 ( 2000-2015 )	論文	于秀玲	吉林大學碩士論文	2017.04			於菟
10	新甲骨文編(增訂本)	新甲骨文編(增訂本)	書籍	劉劍、洪鶴、張新俊	福建人民出版社	2009.05	978-7-211-05853-2	初版	小學堂、陳信良
11	新甲骨文編(增訂本)	新甲骨文編 ( 增訂本 )	書籍	劉劍等	福建人民出版社	2014.12	978-7-211-06715-2	2版	陳信良
12	新見金文字編	新見金文字編	書籍	陳斯鵬、石小力、蘇清芳	福建人民出版社	2012.05	978-7-211-06492-2		陳信良
13	新金文編	新金文編	書籍	董蓮池	作家出版社	2011.10	7-506-35737-1		小學堂、陳信良
14									
15									
16									
17									
18									
19									
20									

「引得市資料來源」<sup>10</sup>表列了文獻版本、索引製作者等訊息

### 三、《引得市》的特點

#### 1、收錄文獻數量豐富：古文字的材質與字體齊全，時代涵蓋久遠

不論是出版的各類書籍，或是海內外博碩士論文等皆有收錄。文獻涵蓋墨跡文字（玉石、竹、木等）與非墨跡文字（甲骨、青銅、石、磚、瓦等），亦包括篆、隸、楷、行、草等字體書法字典相關的文獻。我們知道自有文字以來，字形的演進一直是持續不中斷的，書寫者在傳抄文字當中，刻意或非刻意的留下個人對於該字的理解，同樣的字，若以

<sup>10</sup> 詳見《引得市》Q&A「版本訊息」，或網址：<https://goo.gl/Dh5MFt>

---

不同「筆順」書寫，拉長時間來看就會有明顯的差異。如果我們把時間只縮限在「戰國至漢代」，有些字其實變化的程度並不容易觀察出來。所以，我們必須盡可能地拉長時間，整理補足漢代以後的墨跡文字。而《引得市》目前收錄的文獻也已延伸至唐宋甚至更晚的時代。

## 2、改變研究者查詢工具書的方式（習慣）

大多數研究者還是以紙本逐字查閱工具書，一次只能查詢一本書籍，使用《引得市》的檢索，則一次可檢索數十本文獻書籍，達到所謂「秒查」的速度。若在配合「開卷助理」，還能夠直接瀏覽文獻內容，古文字分析考察研究上更是如虎添翼。

以前不容易執行或者必須花費太多時間的基礎研究工作，因為使用《引得市》，才可以在短時間內順利完成。例如，筆者在戰國時代至漢代的簡牘墨跡文字的研究中，透過資料庫的查詢後，能夠將同一字依照時間（時代分期）順序，形成「文字演進表」就能證明所提出的論點。

通常，古文字演進的考察研究中，筆者會先有一個假設，再以《引得市》快速便利的搜尋方式取得印證，研究者在思考中，隨即又輸入另一字查詢，這種即時的回饋與釋疑，也改變了過去研究的模式。現在，研究者只要拿著能上網的手持裝置（手機、PAD），隨時連上《引得市》就可以查詢古文字，隨想隨得的便利性，勢必影響研究者的研究考察習慣。

## 3、古文字的通盤檢索與缺字查詢輔助

以「古文字缺字資料庫」作為檢索缺字的界面，使用者以文字的某偏旁或部件來查詢，再串連其他資料庫。缺字庫的介紹與解說詳見後文。能夠支援各種文字的缺字查詢並定位連結至書籍文獻，目前網路資料庫中似乎未見。

## 4、索引來源多樣，更新快速、訂正內容多

《引得市》的使用者也多是文獻索引製作者。平時都登入於通訊軟體的「群組」內交換訊息。近年，筆者將目前正在製作或希望製作的文獻索引紀錄在「HackMD」<sup>11</sup>中，以「待辦清單」的方式紀錄目前索引的製作情形。同時，需要製作的文獻也放在群組中請大家「認領」。多人即時的訊息交換，除了索引內容產出快速，也能夠避免重複製作的情形。

索引數位化的製作，優先採取學界最新出版的書籍文獻。索引的製作不像處理全文

---

<sup>11</sup> HackMD 是由國人開發製作，Markdown 線上即時編輯的多人協作筆記。網址：<https://hackmd.io/>

資料庫那樣的耗時繁瑣，因為「字頭索引」內容比較單純，能夠儘速地處理完成，可在短時間內提供研究者檢索使用。一般出版的文字編或字詞辭典等文獻，出版之後，若沒有學者專文討論，則不容易再出版列示其書中錯誤之處，我們製作檢字表的過程中，逐字的與原書對照，故能發現許多原本出版品的錯誤，在筆者部落格或《引得市》將其錯誤與正確的版本列示其中<sup>12</sup>。此外，除了筆者，還有趙瑾昀先生<sup>13</sup>以程式或人力方式逐條檢查內容，每日不斷修正《引得市》的錯誤，讓資料維持在最正確的狀態。大多數的錯誤反饋，我們都能在短時間內完成修正並更新。

## 5、原始文獻數位化後再衍生的新檢索資料

有時候，原作者並沒有製作筆畫檢字表或沒有索引在著作論文中，筆者為了檢索的便利，就自行製作了 PDF 格式筆畫檢字表。如：〈嶽麓秦簡字形比對表〉、〈秦封泥文字編〉、〈魏晉南北朝簡牘文字構形分析表〉等，完成後立即提供自由下載。此外，還有歸類在「簡牘」項目的：《北大藏西漢竹書·蒼頡篇》與《睡虎地秦簡·效律》全文檢索資料庫，屬於原始文獻與筆者原創製作的數位摹本的結合應用。

NO.	名稱	簡號	釋文	圖示	圖版	圖版2:1	釋文頁碼	註釋頁碼	練習本
1	《睡虎地秦簡·效律》	01_00	效	 	0685	0991	0153	0153	01
2	《睡虎地秦簡·效律》	01_01	為(爲)	 	0685	0991	0153	x	01
3	《睡虎地秦簡·效律》	01_02	都	 	0685	0991	0153	x	01
4	《睡虎地秦簡·效律》	01_03	官	 	0685	0991	0153	x	01
5	《睡虎地秦簡·效律》	01_04	及	 	0685	0991	0153	x	01

《睡虎地秦簡·效律》全文檢索資料庫

以上就是《引得市》的基本特色，其中提到的「開卷助理」、「缺字資料庫」等內容，在後段還有詳細的說明，其餘的問題也可以在網站右上方「Q&A」的連結獲得解答。

<sup>12</sup> 《引得市》文獻的部份訂正內容，網址：[www.mebag.com/index/source.asp](http://www.mebag.com/index/source.asp)

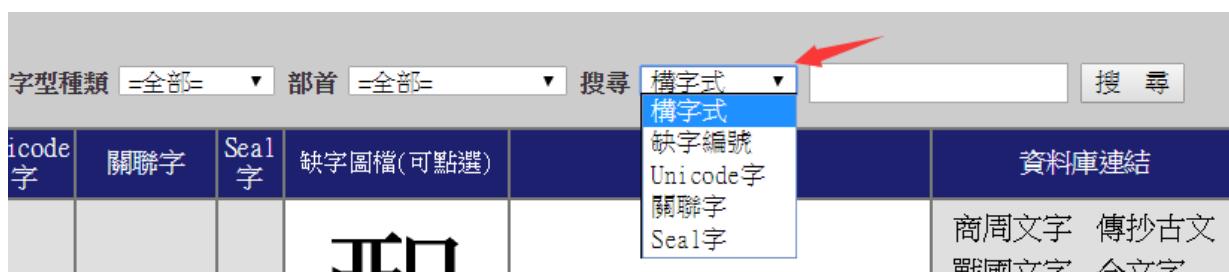
<sup>13</sup> 趙瑾昀先生原為《引得市》的使用者，2014 年與筆者在網路上結識。多年來深耕《引得市》，提供諸多網站發展的建議，資料庫內容與程式修正均出自其手，貢獻卓著，是《引得市》最大的推手之一。

#### 四、《引得市》中的電腦缺字處理

一般的文字查詢檢索在目前來說其實並不稀奇，網路上的文字相關資料庫多如牛毛，但能夠解決或因應「古文字」當中的「缺字」問題則屈指可數。我們所說的「缺字」是指電腦最新的作業系統預設的字型裡無法輸入或顯示的字，這類的字在古文字中佔了相當多的數量，若不能提供檢索，則資料庫的便利性就打了折扣。

例如：《楚文字編》4846 個字例中，缺字就有 2104 字，佔了幾乎是一半的字量。有鑑於此，早在 2004 年筆者在構想建立「先秦璽印分域資料庫」<sup>14</sup>的同時也建構了「印學缺字資料庫」，利用一個「缺字資料庫」來分享支援給日後所建立的各類古文資料庫。以下，先就「古文字缺字資料庫」搜尋欄位分別介紹與說明<sup>15</sup>：

「缺字編號」：	是「古文字缺字資料庫」對缺字所制定編號，依據資料不斷的修正，編號有可能更動的。
「Unicode 字」：	該字相對應的 Unicode 字。(可能是擴張 CDEF，也可能是兼容字)該字實際上已編碼，不過引得市的缺字資料庫為了相容舊式系統依然視為缺字。
「關聯字」：	該缺字在目前使用的文字中的通用字例。
「Seal 字」：	即「古文字缺字資料庫」的所有缺字匯集成的字型檔(seal v2.101.ttf)。使用者安裝此字型檔後，才能在搜尋界面中正確的顯示內容。



「古文字缺字資料庫」界面：使用者可點選「▼」符號，選擇切換檢索的項目

原始缺字資料庫的前身為「印學缺字資料庫」，以「214」個基本的部首作為編號。無法用部首分類的圖形文字則歸類在「215」，「009-001」代表「人」部的第一個字，依此類推。其實，這樣的部首分類並沒有嚴格參考標準，筆者認為，某個缺字應該被分在那一個部首其實並不重要。因為缺字絕大多數是目前傳統的字辭典所沒有收錄的。作為缺字資

<sup>14</sup> 陳信良：〈先秦璽印分域資料庫網站建構研究〉，刊《藝術論文集刊》第二期，47-63 頁，臺北：國立臺灣藝術大學出版編輯委員會，2004 年版，第 50 頁。

<sup>15</sup> 筆者 Blog 拙文：〈20180101「引得市」資料庫與「古文字缺字資料庫」更新及使用說明〉網址：<https://ebag2007.blogspot.tw/2018/01/20180101.html>

料庫的整理，我們只是讓缺字有個分類位置，讓檢索上有個方便性。基本上，使用者還是應該優先以「構字式」來查詢。

多年來累積的「印學缺字資料庫」所包含的缺字已不限於印學方面，擴及所有古文字的領域（甲骨、金文、簡牘等），已順應現況改名為「古文字缺字資料庫」<sup>16</sup>（簡稱「古缺字庫」）。資料庫已製作超過一萬缺字，且依照古文字書籍（出版品）即時的更新增補缺字。目前最多的缺字部首是「邑」部，高達 584 字，若沒有「構字式」的輔助查詢，即使資料庫將該部首全部列表顯示，逐字查閱對照也是相當耗損使用者「眼力」和考驗「記憶」的事。

我們參考「中央研究院漢字部件檢字系統」<sup>17</sup>，利用「構字式」來查詢缺字。例如：輸入「付邑」來檢索，就可檢索到「𠂇」字。除了「構字式」查詢的便利之外，查得的缺字也能知道其「編號」，利用此編號就可再查詢《引得市》內的其他文獻，讓使用者可以知道，缺字在哪一份文獻中的第幾頁（如下圖所示）。「古缺字庫」與《引得市》搭配整合，用「缺字」查詢諸多文獻，一直是過去的研究或資料庫不容易達到的。

每頁筆數	書名	筆劃	文字	頁碼	檢字表頁碼
1	包山楚墓文字全編	7畫	𠂇	0269	0624
2	楚文字編	7畫	𠂇	0401	0915
3	楚系簡帛文字編(增訂本)	7畫	𠂇	0627	x
4	戰國文字編	7畫	𠂇	0429	1155
5	戰國文字編(修訂本)	7畫	𠂇	0429	x
6	戰國古文字典	7畫	𠂇	0392	1575

在古缺字庫中查得缺字「s163-199（𠂇）」，點選「戰國文字」即可跳頁至該資料庫

2013 年起「中央研究院漢字部件檢字系統」已經不再更新，相關內容已轉移在「小學堂文字學資料庫」<sup>18</sup>。在缺字的處理上，小學堂內部有其程式對應的方式，並不是採取因應每年更新的 Unicode 擴展漢字上，這和我們的作法並不相同。而我們在 2014 年起，也和小學堂資訊交流合作。讓時間與人力資源做最有效的發揮。筆者認為「缺字」是古文字字典或文字編是否能夠成功數位化的關鍵。這些文獻透過「古缺字庫」的構字式，也都能夠完整的檢索。

從過去製作過的百餘種文獻來看，戰國時期尤以楚系文字相關字典、文字編中的缺

<sup>16</sup> 古文字缺字資料庫，網址：<http://www.mebag.com/index/quezi/list.asp>

<sup>17</sup> 漢字構形資料庫，網址：<http://cdp.sinica.edu.tw/cdphanzi/>

<sup>18</sup> 小學堂文字學資料庫，網址：<http://xiaoxue.iis.sinica.edu.tw/>

字最多。舉例來說，《戰國古文字典》龐大的缺字數量僅次於四年前處理的《戰國文字編》，為此在幾天內造了一千餘字。之後完成了《楚系簡帛文字編(增訂本)》、《清華大學藏戰國竹簡(陸)》、《清華大學藏戰國竹簡(柒)》等文獻的缺字處理，累積新增了近千字。

近期處理的《古文字譜系疏證》也有三千多個缺字。目前，「古缺字庫」已有 14036 個缺字，達到現階段最高峰。統計後基本漢字與擴展 A、B 區（含）以外的缺字約有 1313 字<sup>19</sup>。只佔了總體缺字的百分之 9.354。可見將近有 90%以上的缺字是目前電腦所沒有的字。因此，這樣的數據比例與缺字內容，作為日後 Unicode<sup>20</sup>新增擴張的字集依據，應該是很有參考的價值。

no	書名	總字列	缺字數	缺字占比	索引完成時間
1	《楚文字編》	4846	2104	43.41	2012.7-2018.1.10
2	《楚系簡帛文字編(增訂本)》	4759	1975	41.50	2017.10.24
3	《古文字譜系疏證》	9693	3535	36.46	2018.2.22
4	《戰國古文字典》	6599	2268	34.36	2015.10.17
5	《上博藏戰國楚竹書字匯》	2399	757	31.55	2013.4.4
6	《戰國文字編(修訂本)》	5963	1633	27.38	2017.9.17
7	《新甲骨文編(增訂本)》	3729	964	25.85	2017.5.6
8	《新金文編》	3015	570	18.90	2012.11.11

表二 各種文獻的缺字比例

每頁筆數	書名	筆畫	文字	頁碼	檢字表頁碼
NO.	書名	筆劃	文字	頁碼	檢字表頁碼
1	楚系簡帛文字編(增訂本)	20畫	饁(錐)	0516	x
2	楚系簡帛文字編(增訂本)	18畫	餓	0508	x
3	楚系簡帛文字編(增訂本)	13畫	餃	0508	x
4	楚系簡帛文字編(增訂本)	14畫	餡	0509	x
5	楚系簡帛文字編(增訂本)	17畫	餉	0509	x
6	楚系簡帛文字編(增訂本)	17畫	餽(僉)	0509	x
7	楚系簡帛文字編(增訂本)	10畫	舍	0510	x
8	楚系簡帛文字編(增訂本)	13畫	奢	0511	x
9	楚系簡帛文字編(增訂本)	9畫	奩	0511	x
10	楚系簡帛文字編(增訂本)	9畫	鉉	0515	x

[ 下一頁 ] [ 最終頁 ]      1 / 198 頁 ( 總計 1975 筆 )

《楚系簡帛文字編(增訂本)》1975 個缺字

<sup>19</sup> 詳細數量分別是：兼容區（6）兼容擴張（8）擴張 C（346）擴張 D（4）擴張 E（906）擴張 F（43）。

<sup>20</sup> Unicode（中文：萬國碼、國際碼、統一碼、單一碼）是電腦科學領域裡的一項業界標準。它對世界上大部分的文字系統進行了整理、編碼，使得電腦可以用更為簡單的方式來呈現和處理文字。維基百科「Unicode」網址：<https://zh.wikipedia.org/wiki/Unicode>。另，Unicode 擴展漢字說明詳見網址：<https://zh.wikipedia.org/wiki/Wikipedia:Unicode%E6%89%A9%E5%B1%95%E6%B1%89%E5%AD%97>

此外，「古缺字庫」累積的一萬多個缺字，結合「小狼毫輸入法」也能有效解決研究者在撰寫文章中缺字的輸入與顯示的煩惱問題。<sup>21</sup>除了《包山楚墓文字全編》與《上博楚簡文字聲系 1-8 冊》之外，《引得市》線上的楚系相關的文字編，幾乎都是筆者所製作，若不是對楚文字的熱愛興趣，在缺字總是佔了近一半內容的情形下，一般人可能不容易持續製作下去。

市面上出版的楚文字相關的文獻如：《上博藏戰國楚竹書字匯》、《楚系簡帛文字編(增訂本)》、《清華大學藏戰國竹簡（壹～柒）》等，都可以在《引得市》上檢索使用。今年年初《楚文字編》所有缺字處理完成的同時，也可以說楚系文字的「缺字」補充應該算是達到一個新階段了。因此，近期我們就公開了最新的缺字字型並提供自由下載。<sup>22</sup>

## 五、《引得市》的「部件檢索」與「開卷助理」

「部件檢索」<sup>23</sup>是由王富國先生所製作，於 2015 年底開始使用，界面可顯示：基本漢字、擴展 A 區、擴展 B 區、擴展 C 區、擴展 D 區、擴展 E 區、擴展 F 區、兼容、兼容擴展字等，能夠完整顯示目前的電腦字型 97247 字。必須注意，使用者如果要能完整顯示，得先要安裝下載「花園字型」<sup>24</sup>或者王富國先生製作的「全宋體」<sup>25</sup>。字型安裝完後，瀏覽器關閉後重新開啟，再前往中國哲學書電子化計劃「字體試驗頁」<sup>26</sup>測試是否能正常顯示擴展 D、E 等所有漢字。如果測試網頁內的字都能夠正常顯示，那麼在「部件檢索」中就能確保可以顯示所有漢字。

「部件檢索」有各種版本的差異，使用前請留意是否是針對某資料庫所設定的專用版本，如果是專用版本，其漢字的檢索範圍就是限定在該資料庫的字頭。基本上，使用者可以在搜尋框內打字進行檢索。若不在搜尋框內打字輸入，也可以點選界面上的鍵盤進行輸入，透過點選「▲」可以隱藏/縮放內建的輸入鍵盤，如：214 部首、部首變體、難輸部件、常用部件、所有部件等，依需求選定「○」進行切換。接著，我們說明「部件

<sup>21</sup> 筆者 blog 拙文：〈20150401 「古文字缺字輸入法」開放免費使用〉，網址：

<https://ebag2007.blogspot.tw/2015/04/20150401.html>

<sup>22</sup> 資料夾中包含：一個說明檔、一個 xlsx 檔，一個 reg 檔，二個字型檔以及一個 zip 壓縮檔。Excel 檔是該版字型檔的清單，列有缺字編號、Seal 字、Unicode 編碼和構字式。壓縮檔，封存有該版所有的缺字圖檔（png 格式），共 13913 個。其他說明詳見網址：<https://zhuanlan.zhihu.com/p/34074690>

<sup>23</sup> 「部件檢索」原作者的製作過程說明：[http://blog.xuite.net/fg\\_wang/twblog/369132335](http://blog.xuite.net/fg_wang/twblog/369132335)。「部件檢索」網址：[http://www.mebag.com/index/word\\_search.htm](http://www.mebag.com/index/word_search.htm)。必須說明，各個資料庫都有其專屬的「部件檢索」

<sup>24</sup> 「花園字型」下載網址：<https://zh-tw.osdn.net/projects/hanazono-font/downloads/68253/hanazono-20170904.zip/>

<sup>25</sup> 見王富國先生 Blog 文章：〈漢字使用環境的建置（一）—— 顯示篇〉網址：<http://fgwang.blogspot.tw/2018/02/blog-post.html>

<sup>26</sup> 「字體試驗頁」網址：<http://ctext.org/font-test-page/zh>

---

檢索」核心版本：0.9.0.8(2017年03月)的基本功能<sup>27</sup>：

- 1.拆分功能**：在搜尋框中，任一文字前增加「\」(斜線符號)，即可以拆解文字，顯示於下方。可選擇複製使用。例：「\部」可拆分為「音(立口) 邑」。有些字可透過這功能，拆分後再別的部件組合成別字，對於我們不知道發音、不知如何輸入的文字，用這種方法就可以解決。
- 2.組字檢索**：在搜尋框中，輸入任一「部件或文字」，下方自動會產生相映的文字。
- 3.點選部件**：此頁面依筆畫排列提供部件，點選部件文字，即顯示於「搜尋框」中，可省略一些必要的輸入文字。部件中也包含「非漢字」的部件，如「小」。
- 4.背景顏色**：文字的背景顏色代表字集如下：

(基本漢字)	e0e0e0 (灰色)
(擴展 A 區)	c0e8ff (淺藍色)
(擴展 B 區)	ffff8d0 (淺黃色)
(擴展 C 區)	d1bada (紫色)
(擴展 D 區)	80ffb0 (綠色)
(擴展 E 區)	ffc0d0 (粉紅色)
(擴展 F 區)	90e0d0 (藍綠色)
(相容)	ffffff (白色)
(補充)	fcc800 (土黃色)

「部件檢索」能將資料庫內所有的「字頭」作詳細的部件拆分，用這樣的方式，研究者可以更深入精確的考察字頭，查詢後點選字即可跳回資料庫的頁面。現在我們已經把這樣的程式應用在《引得市》大部分的資料庫中。例如《北京大學藏西漢竹書·蒼頡篇》<sup>28</sup>全文檢索資料庫。

舉例來說，在搜尋框中輸入「鳥」字，立即可以看到「蒼頡篇」中含有「鳥」部件的所有字例。在「複製模式」沒有勾選的情形下，點選下方的「鳴」字，就會連結到資料庫中（如下圖所示）。我們不需要把資料庫中所有的字都背起來，只要符合其中的部件，就能夠被查詢出來，這就是所謂的「亂槍打鳥、海底撈針」的查詢方式。

如果輸入搜尋的結果超過 99 字，點選搜尋框右側的「►」(三角形符號)就能夠完全把剩餘的字例顯示出來。另外，點選畫面左上方「引得市圖示」，可返回原始資料庫。

---

<sup>27</sup> 筆者 blog 拙文〈2016 年「引得市」小改版說明〉，網址：<https://ebag2007.blogspot.tw/2015/12/2016.html>

<sup>28</sup> 《北京大學藏西漢竹書·蒼頡篇》全文檢索資料庫。網址：  
[http://www.mebag.com/index/beida\\_cangjie/list.asp](http://www.mebag.com/index/beida_cangjie/list.asp)



部件檢索 - 北京大學藏《蒼頡篇》專用版 (1301)

部件：鳥

A set of four small, semi-transparent navigation icons located at the bottom right of the slide. From left to right, they represent: a left arrow, a triangle pointing up, a diagonal line from top-left to bottom-right, and a triangle pointing right.

提示：可利用 “\” 來反向解構漢字，例如 “\犮”。

選項： 包容異體  即時查詢  較大字形  複製模式

字數：總計21字

▲ ■ ● 214部首 ● 部首變體 ● 難輸部件 ● 常用部件 ● 所有部件

» 基本 A 區 B 區 C 區 D 區 E 區 F 區 相容 補充

鳴鳥鳧鳊鳩鳴鳶鴈鴛鴦鴻鵠鷗鷕鷙鷔

鳩 鳩 獭 鳩

核心版本：0.9.0.8(2017年03月)  
發布網址：[部件檢索](#) (WFG製作)

〈部件檢索〉界面 《北京大學藏西漢竹書·蒼頡篇》專用版

每頁筆數	章名	字狀態	搜尋	釋文	排序				
NO.	章名	簡號	釋文	圖示	原寸彩	放大彩	紅外線	釋文	練習本
1	22《鶴鵠章》	68_09	鳴	 	011	035	057	135	17

《北京大學藏西漢竹書·蒼頡篇》全文檢索資料庫

以上的示範是一千多筆字頭的小型資料庫查詢結果。或者將「部件檢索」應用在《說文》中也是相當方便的。首先，進入《引得市》的「說文」項目中，點選右上角的「部件檢索」，在搜尋框中輸入「鳥」字，立即顯示出《說文》含有「鳥」部件的所有文字。如果「複製模式」沒有勾選，點擊文字就可以連結到資料庫中，顯示八種<sup>29</sup>《說文》相關的文獻頁碼。與「部件檢索」的結合，除上述的資料庫，也可使用在《引得市》其他十大資料庫中。<sup>30</sup>對各時期的古文字的部首偏旁、部件等考察研究，相信是很有幫助的。

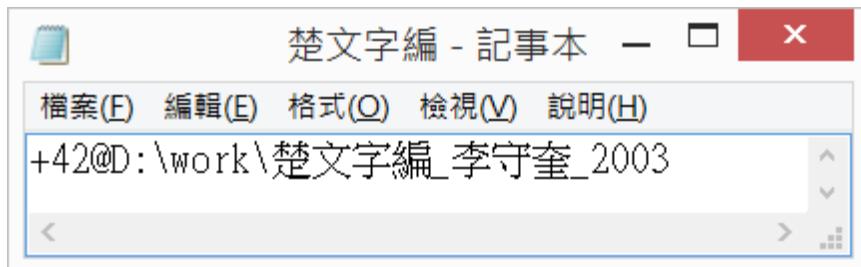
<sup>29</sup> 資料庫界面僅列出簡稱，由左至右依序為：《古文字詁林》、《說文解字考正.》、《說文解字校箋》、《說文解字今釋》、《說文解字約注》、《說文解字集注》、《說文解字探原》、《章太炎說文解字授課筆記》等八種。

<sup>30</sup> 十大資料庫列云如下，字頭數量分別是：商周文字專用版（9954）、戰國文字專用版（16942）、「秦

更值得一提的「部件檢索」作者在 2017 年 6 月開始，提供一個和《引得市》結合應用的重要程式，我們稱為：「開卷助理」(gopage)<sup>31</sup>。這個程式執行時，能讓使用者在《引得市》檢索文獻後，滑鼠游標點選頁碼的同時，讓電腦自動開啟硬碟中的文獻檔案並且正確地跳頁。當然，前提是使用者的電腦中必須自備文獻檔案，否則這個程式不會有任何作用。

NO.	書名
1	楚文字編

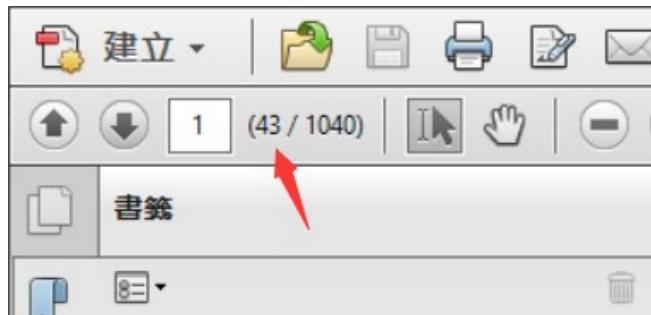
在此之前，「開卷助理」與「文獻檔案」之中，還需要有個連接的橋樑，用文字記事本開啟就可以修改的檔案(副檔名 RMP<sup>32</sup>)，檔案需要和開卷助理的執行檔(GoPage.exe)放在相同路徑下。舉例來說：用於連接《楚文字編》的 RMP 檔一定是「楚文字編.RMP」。檔案名稱必須和《引得市》資料庫中的文獻名稱(書名欄位)一致才行，使用者自己本身文獻的電子檔名則可以自由命名沒有限制。電腦中使用記事本或其他文字編輯軟體開啟「楚文字編. RMP」，內容如下所示：



首先可能會有個疑問，最前面的「+42」數字代表什麼意思？其實這個數字是 Gopage 用來定位第 1 頁的位置。我們只要在 pdf 檔案跳頁匡中輸入 1，右側的數字再減去 1 就是 RMP 內容需要的數字。如果前面沒有其他前言或者目錄內容，那麼就用「+0」。「@」之後的就是 pdf 檔案在使用者電腦檔案的位置路徑。每個使用者依據自己整理的檔案或許會與紙本內容不同。可以自由調整，只要能正確的對應。使用者得依據書籍文獻的不同，修改檔名與內容。其他執行的步驟與教學，在筆者部落格的中都有詳盡的說明。

<sup>31</sup> 「開卷助理」程式與內容是王富國先生設計與製作，授權筆者開放自由下載。「開卷助理」下載網址：<https://goo.gl/RoX2qp>

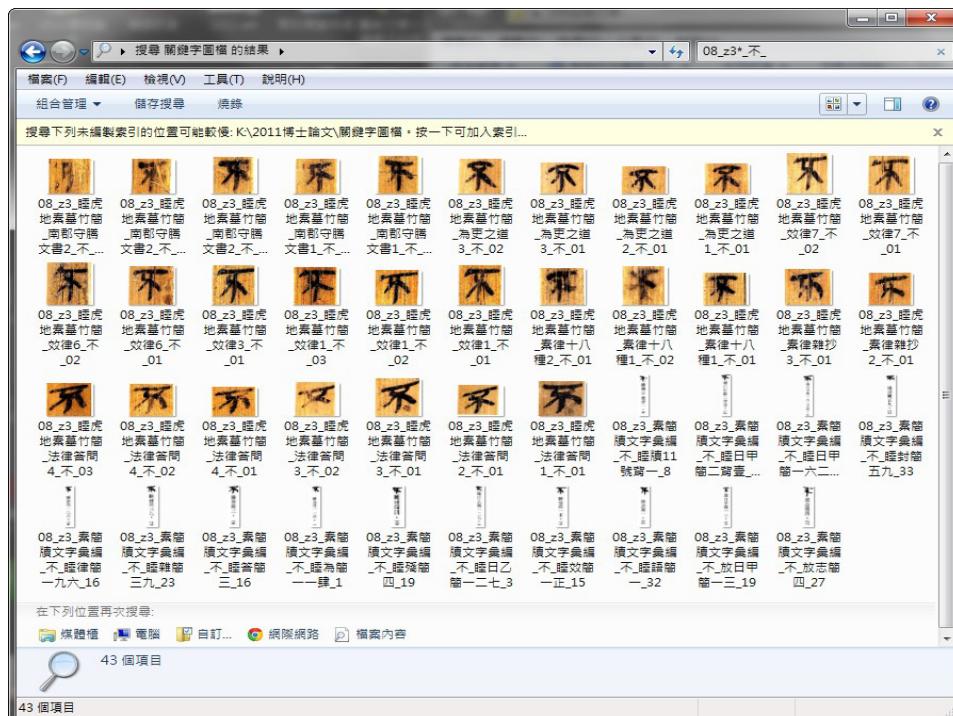
<sup>32</sup> 同註 3，RMP 內容主要是紀錄文獻的定位頁碼和文獻路徑，製作方式如上述，或見 Blog 教學說明。



原本《引得市》提供的僅是書籍頁碼資訊，使用者只能夠依照頁碼翻閱紙本書籍。但是，自從有了「開卷助理」的搭配之後，使用者可以快速的跳頁，瀏覽電腦中的文獻檔案，在漢字檢索與考察中更有效率，由此可知，《引得市》在未來更有存在與發展的價值。

## 六、結語

2012 年拙文〈秦漢墨跡文字演變的考察研究〉提到相關名詞的解釋與考察，強調古文字的分類方法與必要性。又以：「書」「以」「群」「出」「言」「重」「往」「無」「道」「之」「大」「有」「為」「行」「不」「可」「命」，共 17 字為收錄基礎。收錄春秋晚期〈侯馬盟書〉起，至東漢晚期〈東牌樓漢簡〉等墨跡文字，縱括七百多年，期間四十餘種簡牘字跡，超過七千字以上的字例。這是《引得市》建立之前，筆者所作的研究考察方式。經過這一段研究，不再受字體書體名詞的牽絆，也初步的理解文字演進的規律法則。

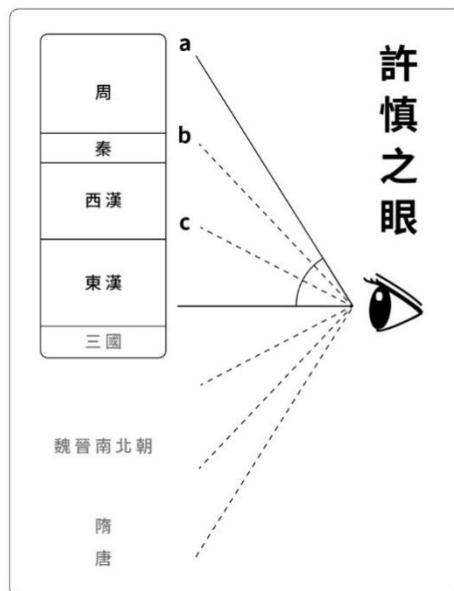


·「不」字的圖檔

一個多月之後，《引得市》建立，有了更多文獻的整合與圖文數據，讓研究立論更加清晰明朗。在 2016 年的〈秦漢墨跡文字的演進考察—以「步」「徙」等字為例〉中，以「步」「徙」「高」等字例對書寫與筆順等關鍵因素作了分析研究。這些都是《引得市》對研究有幫助的實證。

我們試想，生活在東漢中晚期的許慎，流傳下來的傳抄字跡與身邊所看到的文字應該是什麼樣子？過去，我們實在不容易想像。但是近年出土的各時代墨跡文字，讓我們能夠了解後者，圖示中的「角度的高低」代表許慎對古文字認識的深度與資料蒐集的廣闊度，即是他的「眼界角度」。

從目前《說文解字》的版本，我們認為應該是 a 或 b 或 c 那一個角度？無論是那一條線，是絕對不可能在水平線以下的。如果，我們在目前《說文解字》版本中看到了東漢晚期之後才出現的字形結構，我們該如何解釋這個現象？當然，筆者已經有了自己的答案。



許慎之眼

《引得市》第一年草創之初，僅有十七種文獻，現在已經超過二百餘種，成長的速度相當的快。讓研究者有效率的研究，藉此又開發出更為創新的研究方式，是筆者創辦的初衷。研究永遠是以「人」為核心與主體，藉由不同文獻、各層面的資訊，讓研究者自己開發出屬於自己的研究模式，是《引得市》存在與長遠運行的最大價值。

除了筆者獨立整理的索引資料，其他各類的索引資料也是集合眾人之力完成的。為了方便與各地朋友共同製作索引，我們選擇了 WPS Office 作為文獻索引協作的平台。同

---

時多人共同編輯雲端上的文件，讓資料更新的效率又大大的提昇。<sup>33</sup>

「文獻索引的建立」是讓學術研究更為精進的一項重要工作，不只需要對研究有相當的熱忱，也要有獨立冷靜思考的能力。現有的古文字相關的研究圈以外，希望有更多朋友也能加入文獻索引建立的行列，因此，筆者製作了能夠快速學會字頭整理與輸入的說明<sup>34</sup>，只要有興趣、意願就可以立即上手，讓進入索引製作的門檻降到最低。

十四年前，筆者曾說：學界已經投注大量人力、物力在此，可知簡帛方面的資料庫，日後一定是相當的多，惟在璽印、篆刻資料庫的研究使用上，仍有極大的努力空間。羅鳳珠教授說：「中國大陸從 1984 年以來，委託五所大學、五家古籍整理所和一個項目進行試點，結果成效不一。究其原因，一是研究所的負責人或項目負責人未能自己動手運用電腦嘗試古籍整理，這中間大概主要是他們雖是古籍整理的專家，卻不懂電腦而又沒有決心去學習電腦」。<sup>35</sup>

今日，電腦程式等操作界面多有改善，網路的普及，新進的研究人員也熟悉或習慣電腦的使用，因此，古文字相關的網站資料庫，研究單位團體也都發展出各類的資料庫。回頭來看，當時的預測還頗具準確度。

文獻的數位化及各項處理，像是一場一場持續不斷馬拉松，剛衝過終點線，可能又是另一場的起點。筆者個人時間與精力有限的情形下，有些文獻沒有辦法立即完成的都只好暫時擱置。就像《古文字類編（增訂本）》與《新金文編》，雖然早在 2012 年 11 月就公開索引，但是當時的缺字並沒有處理好，字頭完整呈現也是在四年之後。而《楚文字編》也是在六年後才將所有的缺字製作完成。

漢字文獻數位化的前輩莊德明先生曾說：「建立一個好的入口網站當然不是一蹴可幾，除了以往所打下的根基外，往後還須長期的經營。」<sup>36</sup>的確，網站能夠長遠的經營發展是非常重要的。我們也持續構思尋求長期發展的方法，創站以來，所有的索引資料一如往常於《引得市》公開檢索使用，藉此也向原書作者及諸多前賢學者們致以崇高的敬意。<sup>37</sup>

---

<sup>33</sup> WPS Office 是由金山軟件股份有限公司發布的一款辦公軟體，可以應用於辦公軟體最常用的文字編輯、試算表、簡報等功能。覆蓋 Windows、Linux、Android、iOS 等平臺。維基百科「WPS Office」網址：[https://zh.wikipedia.org/wiki/WPS\\_Office](https://zh.wikipedia.org/wiki/WPS_Office)

<sup>34</sup> 筆者 blog 拙文：「引得市」文獻數位化字頭輸入教學懶人包 1.0。網址：  
<https://ebag2007.blogspot.tw/2018/03/10.html>

<sup>35</sup> 陳信良：〈先秦璽印分域資料庫網站建構研究〉，刊：《藝術論文集刊》第二期，47-63 頁，臺北：國立臺灣藝術大學出版編輯委員會，2004 年版，第 49 頁。

<sup>36</sup> 莊德明、鄧賢瑛：〈文字學入口網站的規畫〉，第四屆中國文字學國際學術研討會，山東煙台，2008 年 8 月。

<sup>37</sup> 節錄筆者 blog 拙文：〈20151017《戰國古文字典：戰國文字聲系》索引數位化完成〉，網址：  
<https://ebag2007.blogspot.tw/2015/05/20151017.html>

---

最後，衷心期盼與各界有更多的交流與合作<sup>38</sup>，並給予建議和指教，讓《引得市》更加完善，能為廣大的研究者和民眾提供更好的檢索服務。

## 參考書目

- 陳信良：〈以墨跡文字為主的古文字演進研究方式〉，刊：《中國美術研究》，第 18 輯，南京：東南大學出版社，2016 年版，第 83-90 頁。
- 陳信良：〈秦漢墨跡文字的演進考察—以「步」、「徙」等字為例〉，刊：《藝術研究學報》第九卷第一期，國立臺南大學藝術研究學報編審委員，2016 年版，第 15-38 頁。
- 陳信良：〈文獻的索引數位化製作與應用研究-以《引得市》為例〉，刊：《書畫藝術學刊》第十三期，臺北：國立臺灣藝術大學書畫藝術學系，2012 年版，第 185-212 頁。
- 陳信良：〈秦漢墨跡文字演變的考察研究〉，刊《書畫藝術學刊》第十二期，臺北：國立臺灣藝術大學書畫藝術學系，2012 年版，第 205-278 頁。
- 陳信良：〈先秦璽印分域資料庫網站建構研究〉，刊：《藝術論文集刊》第二期，臺北：國立臺灣藝術大學出版編輯委員會，2004 年版，第 47-63 頁。
- 劉志基：〈古文字偏旁數字平台與數字化環境下的古文字偏旁分析法〉，刊：《中國文字研究》，第 22 輯，上海：上海書店出版社，2015 年版，第 1-9 頁。
- 莊德明・鄧賢瑛：〈文字學入口網站的規畫〉，第四屆中國文字學國際學術研討會，山東煙台，2008 年 8 月。電子版網址：  
<http://cdp.sinica.edu.tw/cdphanzi/documents/T980331.pdf>
- 筆者 blog，網址：<https://ebag2007.blogspot.tw>
- 小學堂文字學資料庫，網址：<http://xiaoxtue.iis.sinica.edu.tw/>
- 王富國先生 blog，網址：<https://fgwang.blogspot.tw/>
- HackMD 多人協作筆記。網址：<https://hackmd.io/>
- WPS 雲文檔。網址：<http://www.wps.cn/>

---

<sup>38</sup> 《引得市》合作提案。網址：[http://hackmd.io/s/HJ\\_qFWLNx](http://hackmd.io/s/HJ_qFWLNx)