

段玉裁《說文解字注》數位內容之設計與建置

羅凡晷*

摘要

許慎《說文解字》開啓了字書以部首歸類的先河，而段玉裁《說文解字注》則是替許書作注的最佳代表，因此想要鑽研文字之學，《說文解字注》是一本必備的案頭書籍。隨著數位時代的來臨，這本著作同樣也被數位化，然而它的數位化進程並沒有走在最前面，其中最大的問題便是中文字碼在處理古籍方面永遠不夠，當 Unicode 字碼出現之後，這個問題才稍有解決的可能。

就筆者所見，與《說文解字》相關的數位化方式約有四類：第一，以書本附光碟方式發行，如：臧克和、王平等人在 2004 年曾編有《說文解字全文檢索》一書，書中附有光碟，但筆者使用上覺得並不方便，因為必須安裝在特定的 OS 平台(微軟簡體中文版)上才能正常開啓，其他 OS 平台並不支援。第二，網際網路的電子全文檢索系統，如：北京師範大學建構了「《說文解字》全文檢索測試版」¹，放置在網際網路上供人使用，使用上頗為方便，可惜的是只有許慎原文而無段注內文。第三，網際網路的圖版檢索系統，如：「中華博物 今古博達網」網站有段玉裁《說文解字注》的圖版檢索²，透過筆畫數即可查找到該字的段注內容，可惜的是只能使用圖版資料，沒有辦法呈現全文電子資料。第四，資料庫公司所建構的《說文解字注》，如：「瀚堂典藏」裡有段玉裁《說文解字注》的全文檢索，資料正確率頗高，可惜的是還有一些中文字無法順利呈現，並且使用這個資料庫的費用不是一般使用者所能負擔的³。

網際網路已邁入知識共享的時代，傳統古籍是人類的公共財，不應被外在的條件(如：價格)限制它的數位化發展，因此，筆者期望建構段玉裁《說文解字注》的全文檢索系統，在使用者需求方面，以滿足段注本全文檢索功能作為主要設計考量；在字碼的處理方面，以 Unicode 為基礎，配合動態組字的方法解決缺字問題，這也正是本資料庫設計的重點所在。

關鍵詞：段玉裁、說文解字注、數位內容設計、全文檢索

* 國立臺灣師範大學國文學系助理教授。

¹ 「《說文解字》全文檢索測試版」，網址：<http://shuowen.chinese99.com/>

² 「中華博物 今古博達網」，網址：http://www.gg-art.com/imgbook/index_b.php?bookid=53

³ 筆者曾建構圖書單位採購，所得結果是一年使用權的費用是新台幣五十萬元左右，這樣的錢不是一般使用者所能負擔與接受的。

一、前言

大專校院中，國文、中文、語教等相關學系，「文字學」幾乎是一門必修的專業科目，在這門必修科目裡，清代學者段玉裁《說文解字注》這本著作也幾乎是任課老師的指定教材或必備參考書。隨著數位時代的來臨，這本著作同樣也被網際網路的使用者所關注，然而它的數位化成果並沒有隨著它的學術價值成正比，主要的原因在於數位化的過程中常常面臨中文字碼不足的問題，以致於數位化工作者多半以外字集造字、圖版掃描等方式來呈現它的數位化成果。隨著 Unicode 編碼方式的不斷進步，就楷體字來說，目前已有七萬多個大字集(Unifonts 5.3 版)可用，也開發出可輸入七萬多字的中文輸入法，因此要處理《說文解字注》這樣一部存在著許多文字異體現象的文字學鉅著，有了更多的可能性。職是之故，筆者擬結合刹那搜尋工坊所開發的 Accelon 3 進行段玉裁《說文解字注》資料庫的建構，讓學子能夠突破傳統學習《說文解字》的困難與障礙，也讓文字學這門學科隨著時代的變遷而有嶄新的面貌。

二、目前所見與《說文解字》相關資料庫簡介

網際網路世界裡，有些網站設計不見得會將網站開張日期置於網頁內容中，就筆者所見所知，最早出現的二個與許慎《說文解字》有關的網頁資料庫，分別是北京師範大學所建置的《說文解字》檢索系統(以提供許慎原文為主，亦有加入大徐本的新附字)與「中華博物 今古博達網」所做的段玉裁《說文解字注》。此後，網際網路與《說文》相關的網站或資料庫，也分別從這兩個方面發展。

(一)許慎《說文解字》的網路面貌

北京師範大學所建置的「《說文解字》全文檢索測試版」⁴，網站上署名是小童所做，從 2004 年 8 月開始提供測試，裡頭除了基本的「關鍵字」查詢，另外還有「高級檢索」，如：「拼音」、「楷字」、「卷」、「正文」、「注」、「重文正文

⁴ 北京師範大學「《說文解字》全文檢索測試版」，網址：<http://shuowen.chinese99.com/>

及注」、「是否新附字」等檢索條件的運用。這個網站主要是以大徐本的《說文解字》為底本，如圖一所示：



(圖一 《說文解字》全文檢索系統首頁)

其中提供網友留言的功能，如果網友有任何意見，都可以在這個地方發言表示自己的意見。如果按下「我要瀏覽」，所出現的畫面如圖二所示：



(圖二 《說文解字》全文檢索系統瀏覽內頁)

由圖二可以看到，這個網站提供了小篆字形及相關的古文、籀文等字形，而字形部分以貼圖的方式呈現，此外並提供楷書、漢語拼音、許慎《說文解字》原文等內容。然而由於《說文解字》一書本身存在著許多的文字異體現象，所以爲了正確呈現資料庫的內容，網站本身有告知使用者應要下載 Ext-B 及相應的超大字符集等。⁵

隨著 Unicode(UTF-8)編碼的發展，多數的文字異體或難字可以順利在網際網路上呈現，因此也出現了其他與《說文解字》相關的網頁資料庫，如 Donald Sturgeon 所建置的「中國哲學書電子化計劃」網站，裡頭便有《說文解字》⁶，如圖三所示：

⁵ 超大字符集，乃目前網路上所流傳的中日韓漢字超大字符集(Super CJK)通用字體支持包(Unicode 5.3 版)，這個字庫包含近 10 萬個標準字符，其中中日韓越通用中文字元 75814 個。因行文之需，詳細介紹，請參見本文「四、楷體難字、小篆、古文、籀文等字形的打字及呈現問題」一節。

⁶ 「中國哲學書電子化計劃」裡的《說文解字》，網址：

<http://chinese.dsturgeon.net/text.pl?node=26160&if=gb>



(圖三 「中國哲學書電子化計劃」網站裡的《說文解字》)

網站所呈現的資料，筆者不清楚所依據的版本為何，有趣的是，資料庫的「卷一」是序文的內容，「卷二」的一開始才是「一部」的「一、元、天、丕、吏」等字，同時提供了全文檢索的功能；另外，這個網站也利用了索引典(或關鍵詞)的連結技術，只要是網站所預設的關鍵詞，按下它的超連結，就可以自動找出該網站所有電子書中具有這個關鍵詞的前後原文，這是此站特別的地方。

另外，有一個網站叫做「《說文解字》在線查詢」⁷，這個網站的最大優點，就是查詢條件相較於其他網站，有了更多的彈性與選項。入口畫面如圖四，在圖右的部分可以看到：它的檢索方式包含了「楷書」、「拼音」、「部首」、「部首筆劃」、「結構類別」、「構件」、「反切」、「卷數」等：



⁷ 「《說文解字》在線查詢」，網址：<http://www.codebit.cn/shuowen/index.php>



(圖四 「《說文解字》在線查詢」首頁)

可惜的是，使用者無法由第一層的入口畫面看到小篆的寫法，必須要按下去某個「楷書」字頭之後，才能看到相應的小篆寫法，如圖五所示：

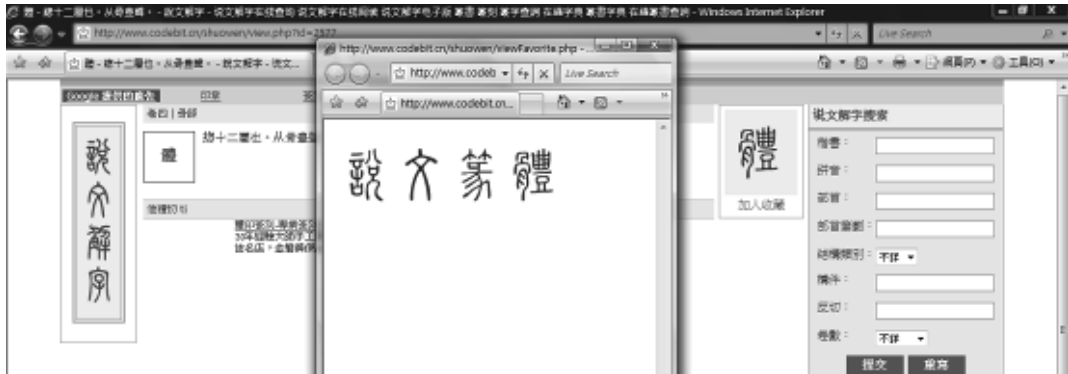


(圖五 「《說文解字》在線查詢」小篆、古文呈現方式)

這個網站有一個它站所無的特色，它能夠將使用者所選的小篆字形「加入收藏」，然後再一次「生成一張大圖」，例如，筆者收藏了「說文篆體」這幾個字，然後再「生成圖片」，結果如圖六所示：



National Chung Hsing University



(圖六 「《說文解字》在線查詢」「生成圖片」的功能展示)

如此可以產出使用者想要看到的小篆字形，然而這只限《說文解字》一書所收的字頭，如果不見於此書的其他楷字，那麼就無法看到相應的小篆字形了。

另外，華東師範大學中國文字研究與應用中心也做了一個「《說文解字》檢索系統」⁸的網頁資料庫，並且取得了很好的成就，它的入口畫面如圖七所示：

⁸ 華東師範大學中國文字研究與應用中心的「《說文解字》檢索系統」，網址：
<http://www.wenzi.cn/shuzi/wenziyuanjiu.asp>。此網站應是該單位所出版的《《說文解字》檢索系統》一書的網路檢索版。



(圖七 「《說文解字》檢索系統」入口畫面)

當筆者按下「《說文解字》檢索系統」右方的「進入檢索」之後，所出現的畫面一開始是亂碼，必須要將網頁編碼改成「簡體中文(GB2312)」才不致於有亂碼問題，編碼改動之後結果如圖八所示：





(圖八 「《說文解字》檢索系統」檢索頁面)

由此圖我們可以看到，目前只提供了「拼音」(漢語拼音)的查詢方式。筆者在輸入欄裡，輸入了「luo2」，當按下「開始查詢」之後，所得到的結果如圖九所示：



(圖九 查詢「luo2」一開始的錯誤編碼頁面)

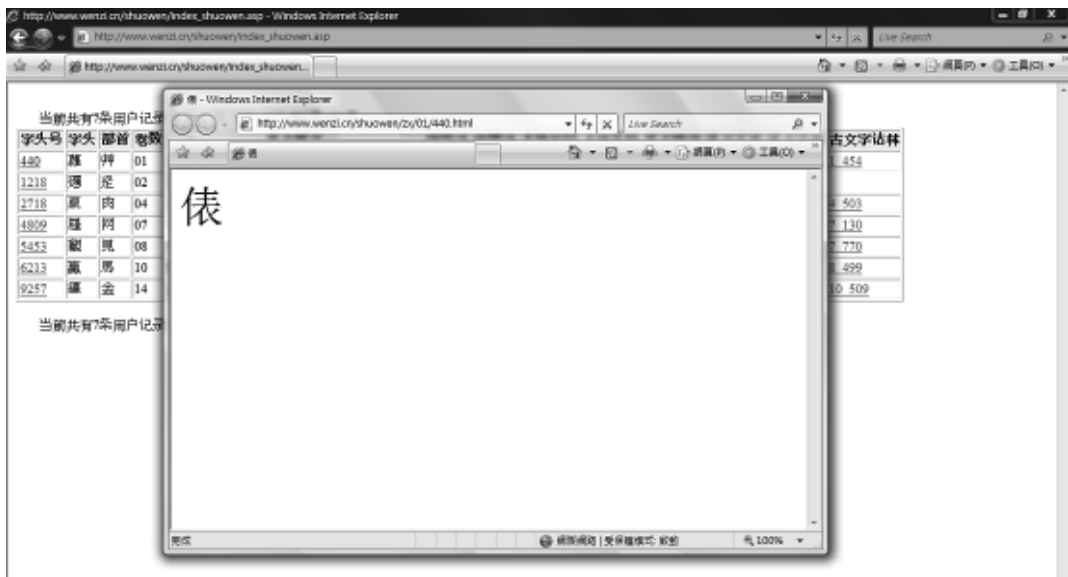
筆者原以為依照前面的使用經驗，應將編碼改成「簡體中文(GB2312)」，結果改動之後還是亂碼，經過一番編碼的不同設定，筆者發現必須將編碼改成「Unicode(UTF-8)」，它才會出現正確的結果，如圖十所示：





(圖十 查詢「luo2」的正確編碼頁面)

根據上圖，我們可以看到，它所提供的查詢結果包含了「字頭號」、「字頭」、「部首」、「卷數」、「說文原文」、「徐鉉注」、「徐鍇注」、「大徐反切」、「大徐頁碼」、「說文段注」、「說文義證」、「說文詁林」、「古文字詁林」等欄位，要製作這些欄位是要花費相當多的時間與工夫的！當我們按下「字頭號」時，會出現小篆的字形，但前題是要安裝華東師範大學所做的小篆字形才能正常看到，否則就會如同筆者的狀況一樣，當按下「440」蘿字時，跳出的畫面變成了如圖十一所示：



(圖十一 查詢「蘿」小篆字型的錯誤頁面)

如果我們按下了「35(說文段注)」，則會出現如圖十二所示的畫面：



(圖十二 打開說文段注超連結的圖版頁面)

可以看到，這個網頁已將相應的段注本全文掃描，提供使用者進行段注本的參照，大大地增加了使用的便利性。

筆者對於華東師範大學的資料庫進行了較為仔細的網頁操作說明，主要的目的是要讓大家看到一件事：目前網頁編碼千百種，就此網站來看，至少用了二種不同的編碼，使用者如果無法正確找到相應的編碼，而逕自以為這個網頁建構有誤，那麼辛苦建構的網站達不到它的實際功效，這樣實在是太可惜了！

(二)段玉裁《說文解字注》的網路面貌

隨著《說文解字》的電子文本資料庫建構風潮，《說文解字注》也不遑多讓，以筆者所見，最早應是「中華博物 今古博達網」所做的《說文解字注》這個網頁資料庫，以楷書筆畫數作為查詢依據，當按下畫面左邊所欲查詢的單字時，右邊視窗則會出現相應的段注內容圖版，在 Unicode(UTF-8)編碼尚未公

佈之前，這個網站提供了使用者極大的便利性，打開了《說文解字注》在網際網路的一扇大門。畫面如圖十三所示：



(圖十三 「中華博物 今古博達網」所做的《說文解字注》頁面)

除了「中華博物 今古博達網」的《說文解字注》以圖版方式呈現之外，華東師範大學中國文字研究與應用中心所建置的「《說文解字》檢索系統」⁹，亦有段注內容的圖檔¹⁰。此外，日本早稻田大學將他們所藏的《說文解字注》(經韻樓藏版)¹¹進行原書全文掃描，這種方式主要是以保存原書圖版樣貌為主，該單位以 html 及 pdf 檔置於網際網路，讓使用者得以觀看該書原貌。

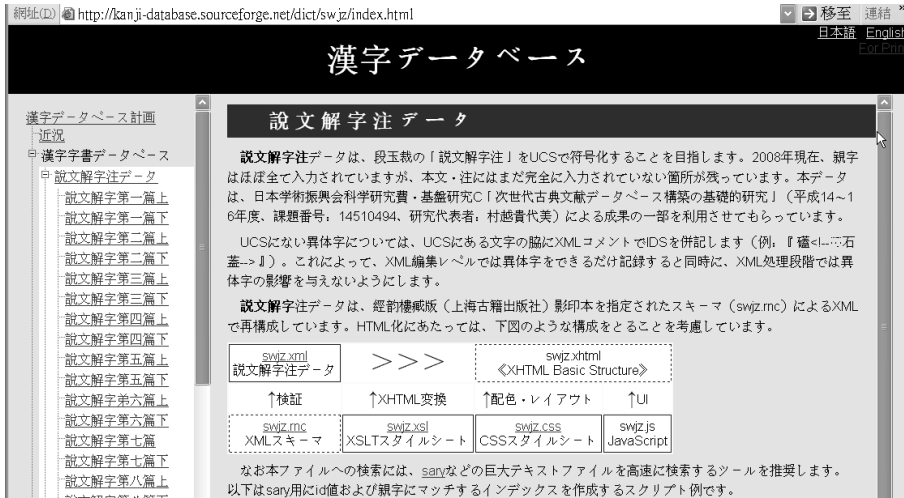
值得注意的是，以日本學者村越貴代美為代表所建置的說文解字注網站¹²，則不再是圖檔的方式，而是以 xml 格式的全文電子檔作為網頁資料庫的建構基礎，如圖十四所示：

⁹ 華東師範大學中國文字研究與應用中心所建置的「《說文解字》檢索系統」，網址：
<http://www.wenzi.cn/shuzi/swjzjs.asp>

¹⁰ 前文已有論述，請參看圖十二所做的說明。

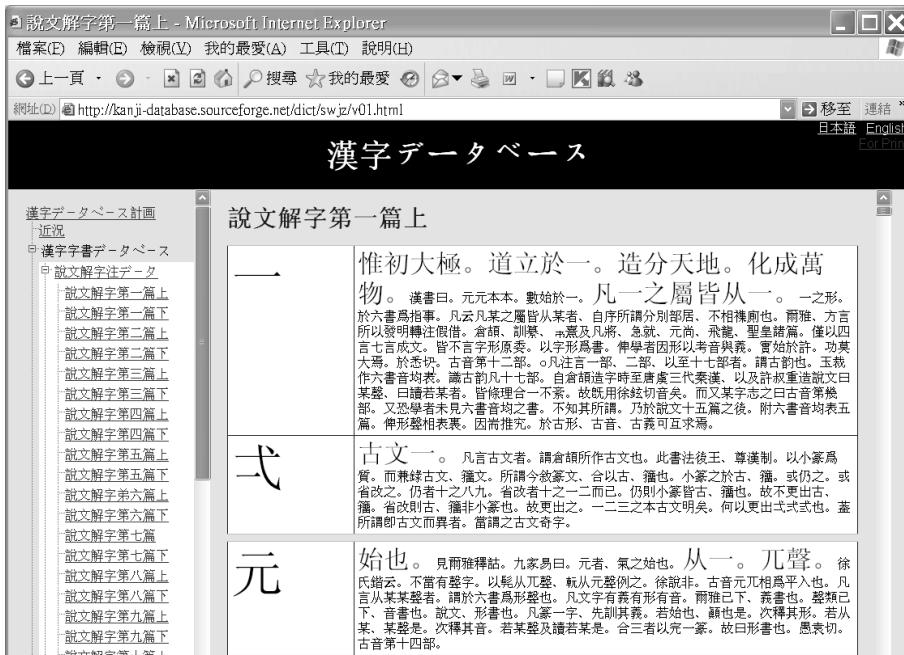
¹¹ 日本早稻田大學所藏《說文解字注》(經韻樓藏版)，網址：
http://archive.wul.waseda.ac.jp/kosho/ho04/ho04_00026/ho04_00026_0001/

¹² 「說文解字注データ」，網址：<http://kanji-database.sourceforge.net/dict/swjz/index.html>



(圖十四 「說文解字注データ」網站入口頁面)

在這個頁面，我們可以看到他們建構網站的基本架構及模式。當我們進一步按下左邊「說文解字第一篇上」時，所得結果如圖十五所示：



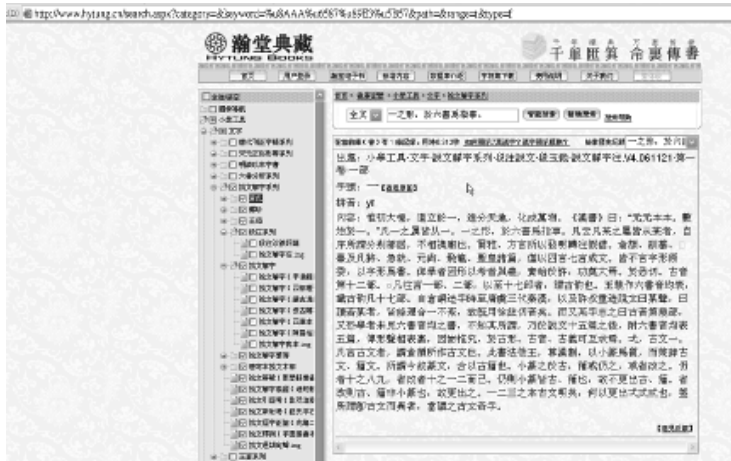
(圖十五 「說文解字注データ」網頁內文)

我們可以看到，此站利用 xsl 樣式表配合 xml 文檔來進行文字的編排。如為字頭，則放大至於表格左側，然而多數的字頭只呈現楷書而非篆體；至於表格右側，則分為大字與小字，其中大字是許慎原文，小字則是段注內文。建構了這樣的一個網站，讓使用者得以進一步利用段注原文進行更多的加值應用，比起圖檔的形式來說，基本上已經克服了大部分的難字問題。當然，這個網站也是在 Unicode 編碼應用之後所建置的。

以上介紹的「中華博物 今古博達網」所做的《說文解字注》、華東師範大學中國文字研究與應用中心所建置的「《說文解字》檢索系統」以及日本學者村越貴代美為代表所建置的說文解字注網站等，使用者在檢索過程當中是不需收費的，這樣比較符合知識共享的網路時代精神，可以讓人類共同累積的知識財產得以迅速的傳播與開展。至於目前市面上資料庫公司所做的中文古籍資料庫，就筆者的瀏覽經驗，有「瀚堂典藏」¹³、「中國基本古籍庫」¹⁴等二個大型資料庫建置了《說文解字注》。其中「瀚堂典藏」可提供個人使用者註冊試用，至於「中國基本古籍庫」則不提供個人使用者試用，必須到有購買這個資料庫的大學圖書館等單位才能使用。由於種種限制，筆者於此僅介紹「瀚堂典藏」的《說文解字注》。下圖是筆者申請了試用的帳號登入之後，選擇「小學工具／文字／說文解字系列」這個選項，並在查詢欄輸入段注內文「一之形，於六書為指事」之後所找到的「全文」條件：

¹³ 「瀚堂典藏」為北京時代瀚堂科技有限公司所建置，相關網址：<http://www.hytung.cn/>。原有台灣代理商，但近日(2008年9月)筆者一直無法順利連結到台灣的網站(<http://www.hytung.com.tw/>)，不知原因為何。

¹⁴ 漢珍數位圖書股份有限公司有代理北京愛如生數字化技術研究中心所建置的「中國基本古籍庫」，相關網址：<http://www.tbmc.com.tw/tbmc2/cdb/intro/Chinese-caozuo.htm>



(圖十六 「瀚堂典藏」查詢「一之形，於六書為指事」所得結果)

根據圖十六，我們可以看到，此系統所提供的查詢結果值包含了「出處」、「字頭」、「拼音」、「內容」等部分，基本上可以正確的顯示段注中的難字，可惜的是，原本應是古文字形的「弌」，只能用相應的楷書寫法處理。使用者如果想要進一步看到原書的面貌，當我們按下「查看原圖」時，所得到的結果如圖十七所示：



(圖十七 「瀚堂典藏」資料庫「查看原圖」結果之一)

在這頁當中還看不到「一」的說文古文，於是必須再按「下一頁」，結果如圖十八所示：



(圖十八 「瀚堂典藏」資料庫「查看原圖」結果之二)

由圖十八當中可以看到原書所呈現的說文古文「弋」的字形，然而回頭看前面所查得的結果，這個部分則是用了相應的楷書「弋」來取代。優點是：如果用楷書來取代，那麼在搜尋的過程中，使用者可以找到這個說文古文的資料；至於缺點，則是使用者如何知道這個字就是說文古文？必須進一步看到段注圖版才能確知這個結果。這個部分則是這個資料庫美中不足之處。

以上介紹了與《說文解字》、《說文解字注》有關的網頁資料庫，除了網頁資料庫的形式之外，亦有隨著圖書所附贈的《說文解字》單機版者。如：臧克和、王平等編了《說文解字全文檢索》一書¹⁵，書後附有光碟版，文獻依據清代孫星衍刻本，是目前流傳最廣的大徐本《說文解字》，能夠查詢全部的說文解字，可惜的是，這個版本只能查看，不能進一步的利用。至於王宏源新勘的《說文解字(現代版)》¹⁶，書後亦附有光碟演示版，但只能查詢說文第四卷及第十一

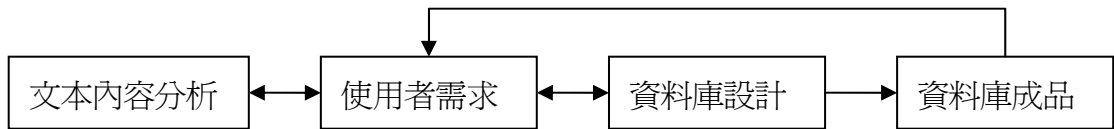
¹⁵ 臧克和、王平等編：《說文解字全文檢索》(廣州：南方日報出版社)，2004年4月第1版。

¹⁶ 王宏源：《說文解字(現代版)》(北京：社會科學文獻出版社)，2005年2月第1版。

卷。以上兩者都只限於許慎的《說文解字》原文，而沒有段注的內容。

三、文本內容分析與資料庫欄位的設計

段玉裁的《說文解字注》裡頭具有許多豐富的資料，如何讓這些資料彰顯出它本身所具有的價值性呢？其中第一個步驟便是進行文本分析的工作，當文本分析完畢，接著便是思考使用者的需求面，了解使用者想要由這個資料庫獲得什麼東西、了解使用者希望在這個資料庫當中看到他想看到的東西等等，這樣的資料庫才有它的存在意義，最後則是進行資料庫的結構(或欄位)設計，配合適當的程式與工具，讓整個資料庫達到它最大的效能。設計流程簡圖如圖十九所示：



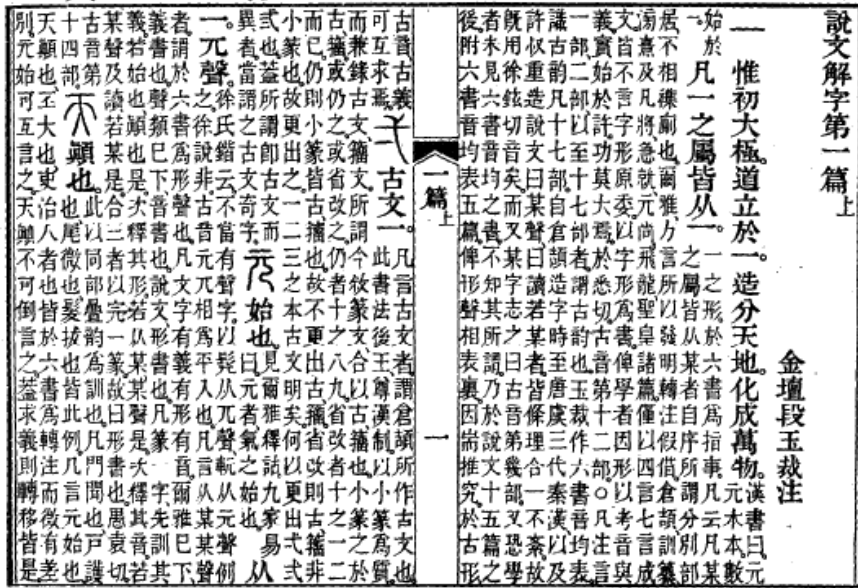
(圖十九 《說文解字注》資料庫設計流程簡圖)

(一)文本內容及使用者需求分析

段玉裁《說文解字注》成書於清嘉慶十二年（西元 1807 年），就流傳的版本來看，主要有嘉慶二十年（西元 1815 年）經韻樓刊本、道光九年（西元 1829 年）學海堂刊《皇清經解》本、同治十一年（西元 1872 年）湖北崇文書局刊本等。本資料庫所使用的文本乃採用嘉慶二十年的經韻樓刊本¹⁷。書前有王念孫的序文，正文版面則如圖二十所示：

國立中興大學

¹⁷ 段玉裁：《說文解字注》（台北：黎明文化事業股份有限公司），1983 年 7 月 10 版。



(圖二十 《說文解字注》內文頁面)

據上可見，前面有書篇名(說文解字第一篇上)及注者(金壇段玉裁注)，接著便是內文。在內文的部分，一開始是小篆字頭，然後是許慎《說文解字》原文，再來是段玉裁的二行注文。以上三者(小篆字頭、許慎原文、段氏注文)基本上是構成此書的主要成分。此外還有一些需要注意的地方，在文本的處理過程當中，如以楷書(楷體字)的繕打角度切入，全文存在著許多的楷書異體狀況，還有大量的非楷書字，如：小篆字頭、「古文」、「籀文」等圖形字，這些都是在進行文本的數位化過程所需面對的。在此時此刻，對於楷書的異體現象，我們可以利用 Ext-A、Ext-B、Ext-C 等解決大部分的問題，少部分無法處理的，則利用剎那搜尋工坊¹⁸作開發的動態組字方式加以處理，而小篆字頭先以小篆字型(如：北師大說文小篆)方式呈現，「古文」、「籀文」等非楷書的形體就暫時以圖版的方式呈現¹⁹。

¹⁸ 「剎那搜尋工坊」，網址：<http://www.ksana.tw>

¹⁹ 其實北師大說文小篆字型裡頭有包含說文古文、籀文字型，但是它的碼位利用了由 Unicode：U+E000 開始之處，如「弋」放在「U+E000」、「𠄎」放在「U+E001」等，在處理上造成字碼對映的困擾。

在使用者需求方面，筆者先進行最基本的全文檢索功能的建置，且擬採用 Accelon 3 程式處理這個問題²⁰。由於目前使用者在網際網路當中只能找到免費的圖版式的《說文解字注》，如果要進一步引用的話，必須自己繕打段注的注文，使用上還是不甚方便；另外有提供全文檢索功能的資料庫，目前只能試用或是付費，對於一般的使用者也不方便，還不如直接查詢紙本原書來得快。因此就目前的現狀來看，提供具備全文檢索功能的《說文解字注》應是可以先暫時解決使用者最基本的需求。

(二) 資料庫設計及成品

設計段玉裁《說文解字注》資料庫的念頭其實很早就 在筆者腦海中誕生了，自從開始進入文字研究領域時，便需要經常翻閱此書，可惜的是，限於電腦科技的相對輔助(主要是中文編碼問題)，直到編碼由 Big5 到 Unicode 的大幅躍進，之後又有 Ext-A、Ext-B、Ext-C 等中文編碼的支援，中文字數大幅增加到七萬多字，傳世文獻的處理相對容易許多，因此筆者在 2006 年 9 月至 2007 年 1 月，配合授課內容之便²¹，以《說文解字注》作為中文打字的輸入文本，請同學們進行中打疑難字的輸入練習，順便看看目前已有的七萬多字超大大字集是否已滿足處理中文疑難字的需求。有趣的是，與筆者大約相同的時間點，大陸及日本方面都有人在進行著同樣的工作，時至今日，筆者在網路上搜尋到段注本的全文電子檔(word 檔)、段注本重編本的電子檔(pdf 檔)、以及段注本全文電子網頁檔(xml 檔)，製作這些文檔的同好者起始點都與筆者相去不遠，可見這個文檔輸入工作已逐漸成熟了。也由於不同電子文本的產生，讓筆者在建構《說文解字注》時多了文本互校的優勢，讓後起的全文電子文本能夠更加的精確，更符合使用者的需求。

有了優良的全文電子文本，如何利用相關的程式設計技術讓它具備更多的加值空間，這也是筆者所探討的一個面向。在此，筆者擬利用剎那搜尋工坊所開發的 Accelon 3 進行資料庫的建構。

²⁰ 關於 Accelon 3 的詳細功能，詳見「剎那搜尋工坊」，網址：<http://www.ksana.tw>

²¹ 筆者此時，分別在台灣師大國文系開設「中文電腦資料處理」及台北大學中文系開設「電腦文書處理」二個班級，修課同學共約有 80 人，每人負責輸入 10 頁左右的《說文解字注》電子文檔。至於輸入的文檔，以 moodle 平台作為中介，供修課同學進行資料庫的整合與建構的練習。

1. 電子文檔編碼及檔案格式處理

由於 Accelon 3 所處理的文檔標記是以 xml 作為架構，因此要將《說文解字注》的全文電子檔改存成「.xml」格式，同時在另存新檔的過程中，筆者利用 emeditor 這套純文字程式加以處理，並設定它的編碼格式為「UTF-16LE」，主要的原因是，段注的全文電子檔有太多的 Ext-A、Ext-B、Ext-C 中才有收錄進來的字，如果沒有注意到編碼格式，那麼文檔必然會有錯亂的狀況出現。為了便於資料庫建構時的除錯檢查，筆者此處僅以《說文解字注》第一篇作為資料庫建構的對象。

2. Accelon3 主架構的建立——shuowen.xml

```
<檔 n="shuowen.xml">
<集 l="書,部,章,原著,說文編號,各體字頭,編目資訊,許慎,段注,註解">說文解字注</集>

<編目資訊 n="原版書名">說文解字注</編目資訊>
<編目資訊 n="版本">版本：清嘉慶二十年歲次乙亥五月刊本。經韻樓臧版</編目資訊>
<編目資訊 n="作者">作者：段玉裁</編目資訊>

<隱><樹 s="*詞" t="書">全文</樹></隱>
<隱><樹 s="*詞" t="章,各體字頭.">篇卷</樹></隱>
</檔>
```

(圖二十一 shuowen.xml)

以上是筆者依照 Accelon 3 程式之需，標記《說文解字注》主架構的 xml 文檔。其中<檔>、<集>、<編目資訊>、<隱>、<樹>等 tag 為 Accelon 3 的內定標記，它的用法及說明請參看「Accelon 愛好者俱樂部」²²，至於「<集 l="書,

²² 「Accelon 愛好者俱樂部」，網址：<http://blog.udn.com/accelon>。又，本資料庫建構時所使用的 Accelon」(剎那搜尋工坊開發，網址：<http://www.ksana.tw/>)，程式端的文檔處理，基本上建構在簡單的 xml 文檔上，至於 DTD 或 XML Schema 的內容，程式開發者並沒有全部公開，筆者只能由相關網頁——如：Accelon 愛好者俱樂部——所提供的步驟進行學習與試作，真有問題再與程式設計師 email

部,章,原著,說文編號,各體字頭,編目資訊,許慎,段注,註解">」乃筆者分析段玉裁《說文解字注》之後所建立的專屬於段注的 tag(即：自定標記),「凡 xml 檔案中的標記,除了內定標記外,所有自定標記必頭在<集>中自行宣告,否則……索引器(indexer.exe)會有錯誤訊息。」²³

3. Accelon3 內文架構標記檔案——00-preface.xml 及 01.xml

Accelon3 的內文架構標記檔案,可以包含一個以上的 xml 檔。因此,在設計《說文解字注》時,筆者先進行文檔的分割,如:書前的王念孫序文單獨建立一個 xml 檔,段注第一篇內文單獨建立一個 xml 檔,其他第二篇、第三篇、…以此類推,這樣的作法有助於文檔的除錯,並且也存在著分工合作的模式,假如未來有更多志同道合之士可以相互合作的話,每個人各分一個部分去做,最後再利用 Accelon3 所具有的文檔整合功能,相信很多事會變得容易許多。

```
<檔 n="00-preface.xml">
```

```
<部>
```

```
<章>王念孫〈說文解字注序〉</章>
```

說文之爲書。以文字而兼聲音訓詁者也。凡許氏形聲讀若。皆與古音相準。或爲古之正音。或爲古之合音。方以類聚。物以羣分。循而攷之。各有條理。不得其遠近分合之故。則或執今音以疑故音。或執古之正音以疑古之合音。而聲音之學晦矣。說文之訓。首列製字之本意。而亦不廢假借。凡言一曰及所引經類多有之。蓋以廣異聞、備多識、而不限於一隅也。不明乎假借之指。則或據說文本字以改書傳假借之字。或據說文引經假借之字以改經之本字。而訓詁之學晦矣。吾友段氏若膺。於古音之條理。察之精、剖之密。嘗爲六書音均表。立十七部以綜核之。因是爲說文注。形聲讀若、一以十七部之遠近分合求之。而聲音之道大明。於許氏之說。正義俗義。知其典要。觀其會通。而引經與今本異者。不以本字廢借字。不以借字易本字。揆諸經義。例以本書。若合符節。而訓詁之道大明。訓詁聲音明而小學明。小學明而經學明。蓋千七

溝通,所以無法在本文清楚的告知讀者 DTD 或 XML Schema 的內容究竟包含了那些。不足之處還請見諒。雖然如此,對筆者而言,Accelon 這個資料庫程式還是充滿著吸引力,而程式設計師也非常熱心幫助筆者解決一些問題,在此一併銘謝。

²³ 「accelon3 資料庫製作手冊快速入門(二)123」,網址：<http://blog.udn.com/accelon/1840055>

百年來無此作矣。若夫辨點畫之正俗。察篆隸之餘省。沾沾自謂得之。而於轉注假借之通例。茫乎未之有聞。是知有文字而不知有聲音訓詁也。其視若膺之學。淺深相去爲何如邪。余交若膺久。知若膺淡。而又皆從事於小學。故敢舉其犖犖大者。以告綴學之士云。嘉慶戊辰五月。高郵王念孫序。

</部>

</檔>

(圖二十二 00-preface.xml)

上面所述內容是序文 xml 檔的全部資料。可以看到這個 xml 檔具有三個 tag，第一個是<檔>，這是屬於 accelon3 本有的內定標記，而<部>、<章>則是筆者的自定標記。

<檔 n="01.xml">

<部>

<章>說文解字第一篇上</章>

<原著>金壇段玉裁注</原著>

<說文編號></說文編號>

<各體字頭>一</各體字頭>

<許慎>惟初大極。道立於一。造分天地。化成萬物。</許慎>

<段注>【段注】漢書曰。元元本本。數始於一。</段注>

<許慎>凡一之屬皆从一。</許慎>

<段注>【段注】一之形。於六書爲指事。凡云凡某之屬皆从某者、自序所謂分別部居、不相禡廁也。爾雅、方言所以發明轉注假借。倉頡、訓纂、滄熹及凡將、急就、元尚、飛龍、聖皇諸篇。僅以四言七言成文。皆不言字形原委。以字形爲書。俾學者因形以考音與義。實始於許。功莫大焉。於悉切。古音第十二部。○凡注言一部、二部、以至十七部者。謂古韻也。玉裁作六書音均表。識古韻凡十七部。自倉頡造字時至唐虞三代秦漢、以及許叔重造說文曰某聲、曰讀若某者。皆條理合一不紊。故既用徐鉉切音矣。而又某字志之曰古音第幾部。又恐學者未見六書音均之書。不知其所謂。乃於說文十五篇之後。附六書音均表五篇。俾形聲相表裏。因

崇推究。於古形、古音、古義可互求焉。</段注>

<各體字頭>弌</各體字頭>

<許慎>古文一。</許慎>

<段注>【段注】凡言古文者。謂倉頡所作古文也。此書法後王、尊漢制。以小篆爲質。而兼錄古文、籀文。所謂今敘篆文、合以古、籀也。小篆之於古、籀。或仍之。或省改之。仍者十之八九。省改者十之一二而已。仍則小篆皆古、籀也。故不更出古、籀。省改則古、籀非小篆也。故更出之。一二三之本古文明矣。何以更出弌弌弌也。蓋所謂即古文而異者。當謂之古文奇字。</段注>

.....

<說文編號> </說文編號>

<各體字頭>毒</各體字頭>

<許慎>厚也。</許慎>

<段注>【段注】毒厚疊韻。三部四部同入也。毒兼善惡之辭。猶祥兼吉凶、臭兼香臭也。易曰。聖人以此毒天下而民從之。列子書曰。享之毒之。皆謂厚民也。毒與竺篤同音通用。微子篇。天毒降災。史記作天篤。</段注>

<許慎>害人之艸。往往而生。从中。</許慎>

<段注>【段注】字義訓厚矣。字形何以从中。蓋製字本意。因害人之艸。往往而生。往往猶歷歷也。其生蕃多則其害尤厚。故字从中。引伸爲凡厚之義。</段注>

<許慎>毒聲。</許慎>

<段注>【段注】毒在一部。毒在三部。合韻至近也。徒沃切。</段注>

<各體字頭>筩</各體字頭>

<許慎>古文毒。从刀管。</許慎>

<段注>【段注】从刀者、刀所以害人也。从管爲聲。管、厚也。讀若篤。筩字、錯本及汗簡、古文四聲韻上从竹不誤。而下譌从副从富。鉉本則竹又誤爲艸矣。古文築作管土。亦管聲。</段注>

.....

</部>

國立中興大學

National Chung Hsing University

```
</檔>
```

(圖二十三 01.xml)

以上 01.xml 文檔內容為《說文解字注》第一篇的部分節錄，一開始是「一」這個字，中間則節選了「毒」這個字²⁴。每個字的基本架構，包含了<說文編號>、<各體字頭>、<許慎>、<段注>等四個 tag，如下所示：

```
<說文編號></說文編號>
<各體字頭>一</各體字頭>
  <許慎>惟初大極。道立於一。造分天地。化成萬物。</許慎>
  <段注>【段注】漢書曰。元元本本。數始於一。</段注>
```

(圖二十四 01.xml 文檔 tag 說明)

其中<說文編號>的 tag 中間不加入參數，<各體字頭>的 tag 中間加入「小篆字形相應的楷書字頭」，<許慎>的 tag 中間則加入許慎《說文解字》的原文，<段注>的 tag 中間則加入段玉裁的注解文字。

另外，爲了讓<各體字頭>這個 tag 裡頭的楷書呈現小篆字形，則必須搭配字型設定的「.txt」文檔，筆者所定的「shuowen_style.txt」檔如下所示：

```
集=細明體,3,$FFCC00
書=細明體,3,$FFCC00
部=細明體,3,$FF3300
章=細明體,2.5,$FFCC00
編目資訊=標楷體,1.5,$FF3300
各體字頭=北師大說文小篆,2,$7F0F0F
許慎=標楷體,1.8,$0F7F0F
```

(圖二十五 shuowen_style.txt)

²⁴ 舉此字的主要目的，在於展示「𠄎富」與「𠄎管土」二組利用動態組字技術的文字處理情形。

從文檔中可以看到，筆者採用了「北師大說文小篆」這個字型，目前這個字型只要安裝了中央研究院所建構的漢字構形資料庫，那麼便會自動將這個字型載入到個人所屬的 fonts 資料夾裡頭，如此可讓 accelon3 所建構的資料庫呈現相應的小篆字形。

4. indexer.exe 及 accelon3.exe 的程式編輯

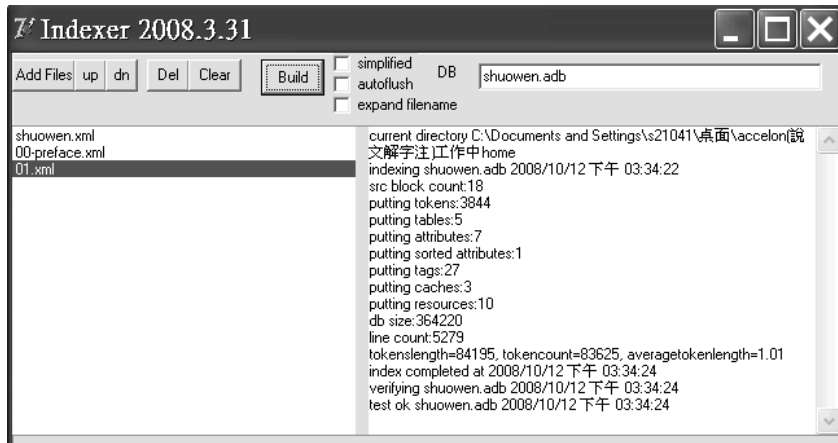
當文檔都已準備齊全，剩下的就是利用剎那搜尋工坊所設計的 indexer.exe 及 accelon3.exe 這二個程式進行編譯工作。筆者在這個地方，要再次的感謝與肯定剎那搜尋工坊將這些程式與技術開放給普羅大眾使用，由於他們的無私精神，讓中文古籍電子化在眾多的形式當中有了另一個有機的開展空間，只要你願意付出你的精力與時間，必然會看到自己努力的成果。

首先，準備好一個「.lst」的文檔，筆者命名為「shuowen.lst」，文檔內容如下所示：

```
shuowen.xml  
00-preface.xml  
01.xml
```

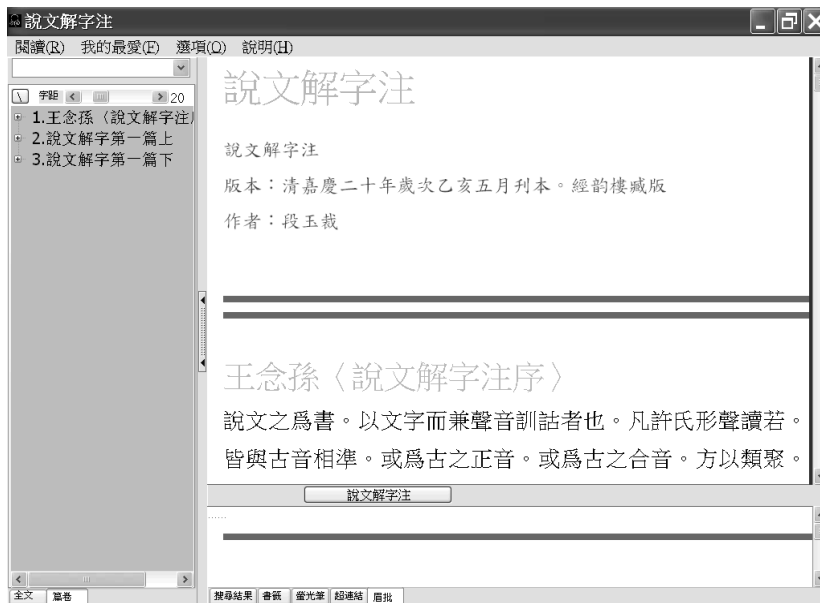
(圖二十六 shuowen.lst)

從這個文檔來看，可以知道，它是將先前筆者所建立的三個 xml 檔案整合在這個檔案裡頭，它的目的是當我們打開 indexer.exe 時，按下「Add Files」時載入「shuowen.lst」，那麼就會出現如下圖左半邊的結果。接著，再按下「Build」，在下圖右半邊便會出現編譯的過程，如果編譯過程中有問題，indexer 會指出問題在什麼地方；如果沒有其他問題，就會呈現如下圖右半邊所示的結果。結束之後，在會同樣的路徑中產出「shuowen.adb」這個電子檔。



(圖二十七 indexer.exe 程式畫面)

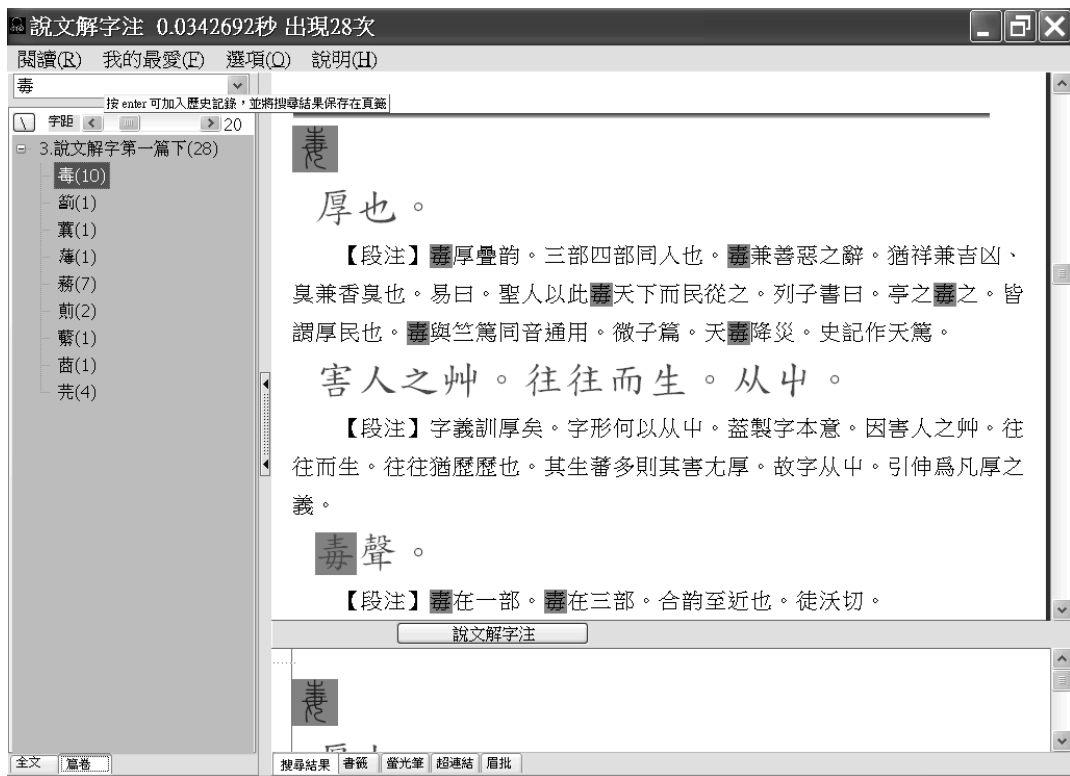
當產出「shuowen.adb」這個電子檔，我們再將這個檔案放到與 accelon3.exe 同一個目錄之中，然後打開 accelon3.exe 這個程式，接著打開「說文解字注」這個資料庫，便會出現下面的結果：



(圖二十八 accelon3.exe 程式執行畫面)

左邊是目錄區，右邊是內文區。在「內文區」的最上面，是筆者在 shuowen.xml 這個檔案裡所定義的版本項。二條灰色的橫線，便是內文的開始。在「目錄區」部分，我們第一個看到的目錄是「王念孫〈說文解字注〉」，這個是由 00-preface.xml 所產生的，至於「說文解字第一篇上」、「說文解字第一篇下」則是由 01.xml 產生的。

此外，如在左半邊上面的查詢欄輸入所要查的內容，如「毒」字，所得結果如圖二十九所示：

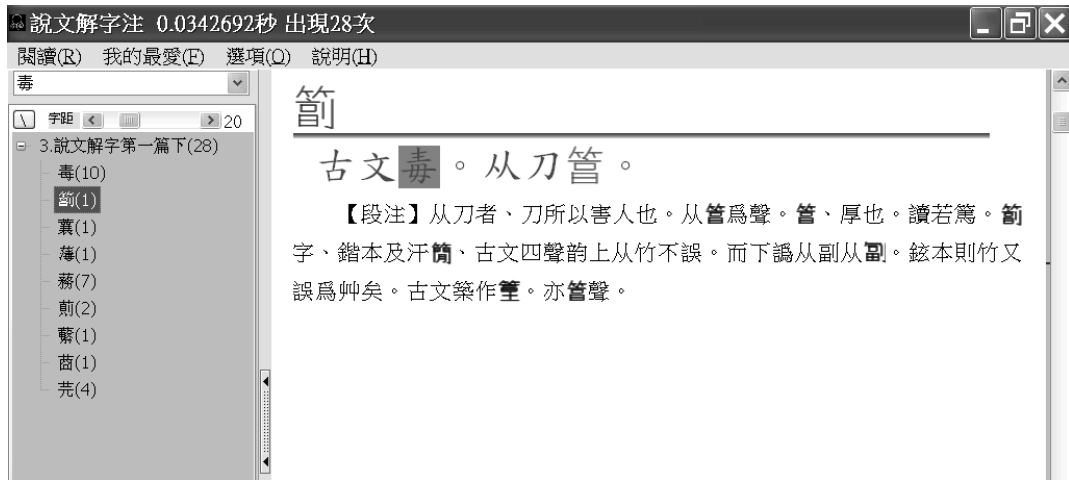


(圖二十九 以 accelon3 查詢「毒」字的搜尋結果)

據上圖二十九可以看到「毒」字在《說文解字注》第一篇當中總共出現二十八次，分別出現在「毒(10次)、箭(1次)、囊(1次)、薄(1次)、蔞(7次)、煎(2次)、繫(1次)、茵(1次)、芫(4次)」等處，同時它的搜尋速度十分快速，這是

accelon3 功能強大之處的其中一個地方。

另外，我們可以看到「毒」這個字的小篆字型可以正常的呈現，如前文所言，這是由於我們在 shuowen_style.txt 裡頭做了字型及大小的設定，讓它可以呈現出我們想要看到的版面樣貌。



(圖三十 accelon3 結合動態組字技術的呈現結果頁面)

至於圖三十，有二個字是在七萬多字的超大字符集之外的，分別是「毒」及「管」字，如果打開 01.xml 文檔(可參見本文上面所引之 01.xml 處)，在這個地方則是用了動態組字的技術，分別是「[富]」與「[管土]」，目前這個動態組字技術乃利那搜尋工坊有別於其他之處，也由於這個動態組字技術，讓中文字有了不再缺字的可能性存在²⁵。

四、楷體難字、小篆、古文、籀文等字形的打字及呈現問題

前文所言七萬多字的超大字符集，乃目前網路上所流傳的中日韓漢字超大字符集(Super CJK)通用字體支持包(Unicode 5.3 版)，根據說明，這個字庫包含近 10 萬個標準字符，其中中日韓越通用中文字元 75814 個，如表一所示：

²⁵ 原則上以楷體字來說是沒有太大的問題的，然而就圖形字而言則沒有辦法進行處理。

CJK 基本	[4E00-9FFF]	20992 碼位	實際 20924 字
CJK 擴展 A(Ext-A)	[3400-4DBF]	6592 碼位	實際 6582 字
CJK 擴展 B(Ext-B)	[20000-2A6DF]	42720 碼位	實際 42711 字
CJK 擴展 C(Ext-C)	[2A700-2BA7F]	4224 碼位	實際 4219 字
CJK 相容擴展	[2F800-2FA1F]	544 碼位	實際 542 字
CJK 部首擴展	[2E80-2EFF]	128 碼位	實際 115 字
CJK 康熙字典部首	[2F00-2FDF]	224 碼位	實際 214 字
CJK 筆劃	[31C0-31EF]	48 碼位	實際 36 字
CJK 相容	[F900-FAFF]	512 個碼位	實際 474 字

(表一 中日韓越通用中文字元碼位對照表)

利用這個超大字符集，配合相應的中文輸入法，如：第五代倉頡輸入法²⁶、海峰五筆²⁷等，便可顯示和輸入七萬多漢字。雖然如此，這些字當中存在著許多的重碼問題，例如：網友謝振斌等人曾將他所看到的重碼字列舉出來²⁸，又如網友阿勤有一篇文章〈从《劍劍劍劍劍劍劍》、《戶戶戶》、《兌兌》等等收字談起〉²⁹，我們光看他的文章標題，就可以看到楷書異體的眾多面貌，這些楷體字型的出現，其實就是對於傳世刻本字體的鏡射現象，也由於這種如實的表現楷體面貌，讓現在的電腦漢字數量永遠追不上使用者的需求。然而反過來說，這些超大字符集的出現，也正是想努力追上使用者的電腦用字需求呀！這當中如何取得平衡，實在是一門大學問！

本資料庫在建構的過程當中，也遇到了小篆字形呈現的問題。就目前學術界、商業上等所使用的小篆字體，其實各有優缺點，如網友李小狼比較了「方

²⁶ 「倉頡之友」，網址：<http://www.chinesecj.com/newsoftware/index3.php?Type=1>

²⁷ 「海峰五筆」，網址：<http://okuc.net/software/>

²⁸ 「電腦中的重碼字」，網址：<http://www.pkucn.com/viewthread.php?tid=164003&extra=page%3D1>

²⁹ 阿勤：〈从《劍劍劍劍劍劍劍》、《戶戶戶》、《兌兌》等等收字談起〉，網址：

<http://www.pkucn.com/viewthread.php?tid=225427&extra=page%3D2>

正小篆體」、「北師大說文小篆」、「全字庫說文解字」³⁰、「漢儀篆書繁」、「中國龍金石篆」等五種篆體字形，發現每種篆體都不盡人意。³¹筆者改繪成表格，如表二所示：

字體名稱	優點	缺點
方正小篆體	最有毛筆味，書法感覺最好，有些部件的寫法(如升、斗、尸、并、欠、異等)比說文更正確，「儿」、「乃」等部件較說文統一。	有些字寫法有誤(如鄉、饗、儿、丰、少、壬等)。「大」部件仍不很統一。另外有時書法式的寫法，若用作字典等標準中未必最合宜(在下看到有些古文字字典，已用方正小篆作其篆體之顯示)。有些自己補的字寫法有問題。
北師大說文小篆	字形依說文小篆，也不會擅加字。	沒英數符數，字體美觀度未是最高(主要在於筆畫較方正、漢儀的生硬)。部分字的風格不統一(在說文裏是古文、重文的話，會用不同書風，這字體依說文使用不同書風，以致風格不統一)。說文字樣不太正確(如升、斗、欠、尸等)或有不統一的(如「儿」、「大」、「乃」等)，仍會依隨。缺字太多。
全字庫說文解字	以說文解字注的篆形製作，使我們除大小徐本說文的字樣外，還有多一個參考。	字體質素極劣，每字的大小、粗細不一，連字寬也成問題，以致有些字總是疊着別的字。描邊質素差，若經人為地修正字寬後，開發成「古印體篆書」似乎較好……

³⁰ 「全字庫說文解字」，網址：<http://www.cns11643.gov.tw/cns11643/images/ttf/ebas927.ttf>

³¹ 詳見「北大中文論壇／中文信息處理／正篆體字形？」，網址：

<http://www.pkucn.com/viewthread.php?tid=219185&extra=page%3D1>

漢儀篆書繁	字樣主要依照說文，比較標準一些。而筆畫自己組出來，組細均勻，字形美觀。	只有 GB 碼版，不能支援繁體字，屬「假繁體」字形，使用時亦無可避免碰上「簡體別字」之毒，在「簡體別字」中只收錄了其一。有較多自己補的字寫法有問題。說文裏的錯誤或不統一，通常仍保留下來（雖也有一些修正）。
中國龍金石篆	Big-5 字型，收字較齊全。	字形的粗細美觀不及北師大。有些自己補的字寫法有問題。說文裏的錯誤或不統一，仍保留下來。

(表二 小篆字體優缺表)

(資料來源：<http://www.pkucn.com/viewthread.php?tid=219185&extra=page%3D1>)

除了網友所列這些小篆字體，筆者尚看過逢甲大學中文系宋建華教授所製作的「說文標篆體」，這套篆文乃宋教授以個人之力花了一年九個月的時間製作（1997年10月~1999年6月），至於篆文的製作底稿，主要是參考清乾嘉時說文兼篆書家沙青巖氏所撰《說文大字典》，並依實際需要加以修改。此外，這套篆文實際字數近一萬五千字，遠超過《說文解字》字數，其目的乃兼顧其他之用途（如美工，一般文件等），因此在字形結構上，完全依據楷書結構組合。另外，華東師範大學中國文字研究與應用中心也製作了一套說文字型³²，篆文字形出自北宋校本《說文》真本汲古閣版，可惜的是，這套字型建立在簡體編碼，目前尚未支援 Unicode 編碼，使用上存在著一定的困擾。至於《說文解字注》當中的「古文」、「籀文」字型，則沒有另外為它們做一套字型，多半是包含在學術界所製作的小篆字體裡，到底要包含在裡頭還是另外做一套字型，其實這也是值得重視的一個問題。³³

綜上所述，雖然現在的 Unicode 四位元組編碼方式搭配配合的中文輸入法

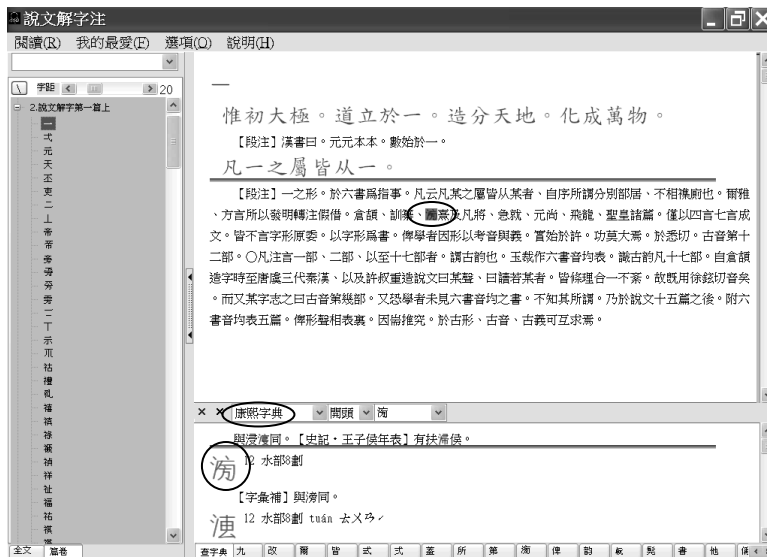
³² 「《說文解字》檢索系統，網址：<http://www.wenzi.cn/shuzi/shuowen/shuowenshuoming.HTM>。這個網頁提供了小篆字型的下載，可參看。

³³ 此處筆者只將小篆字型的現狀提出，至於深入的分析，有待來文加以探討。

已可處理 7 萬多字，然而這僅限於楷書字體，對於段玉裁《說文解字注》這本著作裡還有小篆、古文、籀文等非楷書字體，仍無法有效的支援並進行檢索工作。而筆者於此至少解決了小篆字型搜尋問題，至於古文、籀文等目前只能用圖形加以處理，如果要解決搜尋問題，則必須有一套獨立的字體方可實現。這也是筆者未來可以努力的地方。

五、結語

本文嘗試賦予段玉裁《說文解字注》新的時代生命，以電子文本的方式配合利那搜尋工坊所開發的 accelon3 這個小而美且功能強大的資料庫軟體進行設計與建置，成功地建立了一個「《說文解字注》資料庫」，單獨使用有它的可看性，如結合其他資料庫則更顯出它的功能與特色，例如：筆者將《說文解字注》與王志攀先生所整理的《康熙字典》置於同一個目錄底下，當執行 accelon3 程式並打開《說文解字注》之後，如用滑鼠在畫面右半邊圈選一個「滂」字，那麼在右半邊的底下便會出現《康熙字典》中「滂」字的解釋。這是一個十分有用的功能，可讓使用者進行跨資料庫的查詢與檢索。如圖三十一所示：



(圖三十一 accelon3 跨資料庫的查詢與檢索結果頁面)

此外，將七萬多字超大字符集無法顯示的楷體字以動態組字方式在資料庫中呈現正確的字型，這是本資料庫特出之處；至於在資料庫中呈現正確的小篆字體，並且在全文檢索時也能夠搜尋到小篆字體，這也是本資料庫與眾不同之處。

為了便於讀者比較，筆者以「建置者、說文版本歸類、編碼、小篆字形、基本檢索方式」等項目進行了相關網站的比較，如下表三所示：

網站名稱	建置者	說文版本歸類	編碼	小篆字形	基本檢索方式	網址
《說文解字》全文檢索測試版	1.北京師範大學 2.小童	大徐本	utf-8	有，以圖版方式呈現	以「關鍵詞」的方式進行字頭及內文的全文檢索	http://shuowen.chinese99.com/
《說文解字》	Donald Sturgeon	不明(只有許慎的原文)	utf-8	無	全文檢索	http://chinese.dsturgeon.net/text.pl?node=26160&if=gb
《說文解字》在線查詢		不明(只有許慎的原文)	utf-8	有，但在第二層呈現	包含了楷書、拼音、部首、部首筆劃、結構類別、構件、反切、卷數等	http://www.codebit.cn/shuowen/index.php
《說文解字》檢索系統	華東師範大學中國文字研究與應用中心	以許慎原文為呈現資料，其他欄位則有不同的書籍資料	編碼不統一，有GB2312、utf-8	有，必須安裝該單位所製作的小篆字型檔	只提供漢語拼音	http://www.wenzi.cn/shuzi/wenziyuanju.asp

「中華博物 今古博達 網」	不明	段注本	big5	有，圖版	按照筆畫 順序檢索	http://www.gg-art.com/imgbook/index_b.php?bookid=53
「說文解字 注データ」	以日本學者村越貴代美為代表	段注本(經韵樓臧版(上海古籍出版社))	utf-8	無	無(只按說文卷次排序)	http://kanji-database.sourceforge.net/dict/swjz/index.html
「瀚堂典藏」的《說文解字注》	北京時代瀚堂科技有限公司	段注本	utf-8	有，但在所附圖版的超連結裡	支援出處、字頭、拼音、內容等檢索條件	http://www.hytung.cn/
《說文解字注》資料庫	筆者	段注本(經韵樓臧版(台北：黎明))	utf-8	有，可以任何一套小篆字型呈現(目前以北京師大所做的小篆字型為主)	目前以字頭、全文等作為查詢的主要條件。	目前以單機版為主，未來會將完成的電子書以網頁方式呈現(Accel on3 支援網頁及動態組字的呈現)

(表三 與說文解字相關網站及資料庫比較表)

當然本資料庫還有未竟之處，例如在電子文本的處理過程中，筆者請八十位左右的同學幫忙繕打，因此電子文本的來源乃出自於眾人之手，且由於同學們仍在學習如何輸入楷體疑難字階段，加上每個人處理能力不見得相同(首先，要正確繕打出七萬多字的超大字符集，目前只有第五代倉頡輸入法及海峰五筆輸入法可供使用，因此文稿繕打者必須先學會其中一種；其次，當學會之後又必須要能夠正確拆碼，而這個是需要一段較長的熟練時間；第三，文稿繕打者必須細心繕打及仔細校正，才有品質好的電子文本；第四，有些字是打不出來的，目前只能用圖檔方式呈現，而圖檔的後置處理又必須花費不少的時間。)

雖然現在已有其他不同的電子文本可供互相參校，然而筆者在進行統整與校正的過程還是花了不少的時間。筆者至此深感一個人的能力是有限的，如何能夠集眾人之力完成這項工作，則是中文古籍電子化一項重大的課題。另外，在小篆字體(或加上古文、籀文等字體)的處理上，也是一項難題，到底小篆字體要如實的呈現某個版本的篆形呢？還是將近人研究小篆字體的成果表現在實際呈現之中呢(例如：段氏改篆問題是否要加以處理)？

綜上所述，目前筆者所建置的「《說文解字注》資料庫」提供了全文檢索的功能，而這個功能至少滿足了使用者最基本的查詢需求，例如：使用者想要了解段玉裁引用「禮記」的地方，那麼在查詢欄輸入「禮記」二字，便可在《說文解字注》第一篇看到這個詞共出現二十一次。至於較為進階的檢索應用方式，有賴於使用者依照個人的需求加以設定³⁴，求得最佳的應用結果！

參考文獻

一、書目及期刊論文

王宏源，《說文解字(現代版)》(北京：社會科學文獻出版社)，2005年2月第1版。

段玉裁，《說文解字注》(台北：黎明文化事業股份有限公司)，1983年7月10版。

臧克和、王平等編，《說文解字全文檢索》(廣州：南方日報出版社)，2004年4月第1版。

二、網站或網頁文章

「《說文解字》全文檢索測試版」，網址：<http://shuowen.chinese99.com/>

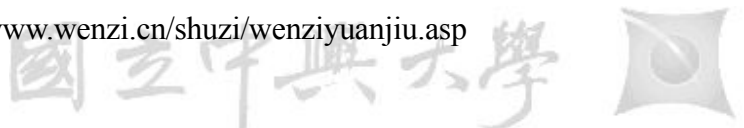
「《說文解字》在線查詢」，網址：<http://www.codebit.cn/shuowen/index.php>

「《說文解字》檢索系統」，網址：

<http://www.wenzi.cn/shuzi/shuowen/shuowenshuoming.HTM>。

³⁴ 進一步的查詢及操作方式，可參看「accelon 操作說明」，網址：<http://www.ksana.tw/help/>

- 「accelon3 資料庫製作手冊快速入門(二)123」, 網址：
<http://blog.udn.com/accelon/1840055>
- 「Accelon 愛好者俱樂部」, 網址：<http://blog.udn.com/accelon>
- 「accelon 操作說明」, 網址：<http://www.ksana.tw/help/>
- 「中國哲學書電子化計劃」《說文解字》, 網址：
<http://chinese.dsturgeon.net/text.pl?node=26160&if=gb>
- 「中華博物今古博達網」所做的《說文解字注》, 網址：
http://www.gg-art.com/imgbook/index_b.php?bookid=53
- 「北大中文論壇／中文信息處理／正篆體字形？」, 網址：
<http://www.pkucn.com/viewthread.php?tid=219185&extra=page%3D1>
- 「全字庫說文解字」, 網址：
<http://www.cns11643.gov.tw/cns11643/images/ttf/ebas927.ttf>
- 「倉頡之友」, 網址：<http://www.chinesecj.com/newsoftware/index3.php?Type=1>
- 「海峰五筆」, 網址：<http://okuc.net/software/>
- 「電腦中的重碼字」, 網址：
<http://www.pkucn.com/viewthread.php?tid=164003&extra=page%3D1>
- 「說文解字注データ」, 網址：
<http://kanji-database.sourceforge.net/dict/swjz/index.html>
- 「瀚堂典藏」, 網址：<http://www.hytung.cn/>
- 日本早稻田大學所藏《說文解字注》(經韻樓藏版), 網址：
http://archive.wul.waseda.ac.jp/kosho/ho04/ho04_00026/ho04_00026_0001/
- 北京愛如生數字化技術研究中心「中國基本古籍庫」, 網址：
<http://www.tbmc.com.tw/tbmc2/cdb/intro/Chinese-caozuo.htm>
- 阿勤：〈从《劍劍劍劍劍劍劍》、《戶戶戶》、《兌兌》等等收字談起〉, 網址：
<http://www.pkucn.com/viewthread.php?tid=225427&extra=page%3D2>
- 華東師範大學中國文字研究與應用中心「《說文解字》檢索系統」, 網址：
<http://www.wenzi.cn/shuzi/wenziyuanjiu.asp>



Design and Build Digital Content with *Notes of Shuo Wen Jie Zi* Composed by Duan Yu-cai

Fan-chen Lo*

Abstract

The book of *Shuo Wen Jie Zi*(說文解字), written by Xu-shen(許慎), is the first book edited using traditionally recognized components of Chinese characters. *Notes of Shuo Wen Jie Zi*(說文解字注), written by Duan Yu-cai, is the best book of annotations. With the arrival of the digital age, *Shuo Wen Jie Zi* and *Notes of Shuo Wen Jie Zi* are created digitally; however, they have not been successfully transmitted. One major problem is the insufficient amount of the Chinese characters code. However, the problem can be solved with the appearance of the Unicode.

Currently, we can find four digital contents with *Shuo Wen Jie Zi* on the World Wide Web. First, CD Rom books, such as *Shuo Wen Jie Zi Full-Text Search* written by Zang Kehe and Wang Ping. However, I feel this CD Rom is insufficient because it must be setup in a special OS (just like windows XP (simple Chinese)), and it does not support another OS. Second, full-text search engines on the World Wide Web, such as *The Test Version of Shuo Wen Jie Zi Full-text Search* created by Xiao Tong(小童) can be used by anyone accessing the World Wide Web and is very helpful. However, it only contains the book of Xu Shen's words and does not contain the book of Duan Yu-cai's words. Third, the pictures search engine system on the World Wide Web, such as the web-site of 'Zhong Hua Bo Wu', can be searched

* Assistant Professor, Department of Chinese, College of Liberal Arts, National Taiwan Normal University

using stroke number, but it cannot present text data. Fourth, *Notes of Shuo Wen Jie Zi*'s database by a Database Corp., such as Hytung corp., has high accuracy, but there are some problems with Chinese characters code because of the display. In addition, this database is costly to use.

It's time to enter an era of knowledge-share-web. The original books are humanity's wealth, and they can not be locked away by external conditions and the limitations or digital development. Thus, I hope contributing a full-text searching engine system with the book of *Notes of Shuo Wen Jie Zi* will be useful. The goal is that users can search for any kind of words in the system. In the Chinese characters code, I designed this system with Unicode, and solved the word loss problem by dynamic-composing-word(動態組字).

Keywords: Duan Yu-cai, *Notes of Shuo Wen Jie Zi*, Design of Digital Content, Full-text Search



National Chung Hsing University