

文字學入口網站的規畫

莊德明 中央研究院資訊科學研究所研究助技師
鄧賢瑛 中央研究院資訊科學研究所研究助理

摘 要

漢字有形、音、義三要素。在字形方面，中研院資訊所文獻處理實驗已於 2008 年 5 月推出漢字構形資料庫 2.52 版，這個版本收錄古今漢字 116,790 個；其中楷書字形 62,942 個，小篆及重文 11,100 個，金文 21,413 個，楚系簡帛文字 19,138 個，甲骨文 2,197 個；另收《漢語大字典》異體字表 12,208 組。我們已利用漢字構形資料庫開發漢字部件檢字系統，並用來解決缺字問題。經過這些年來的驗證，這個缺字問題的解決方案，基本上是可行的。目前，「數位典藏與數位學習國家型科技計畫」至少已有 23 個資料庫採用漢字構形資料庫來解決缺字問題。

雖然歷經十五年的開發，漢字構形資料庫不論在資料庫本身或推廣方面，仍有很大的改善空間：一、漢字有形、音、義三要素，漢字構形資料庫目前只著重在字形知識的整理，尚未涉及字音與字義；建立一個形、音、義俱備的漢字知識庫，仍是我們長遠的目標。二、漢字構形資料庫的缺字解決方案，採用單位仍侷限於中研院及「數位典藏與數位學習國家型科技計畫」，仍須再推廣。三、部件檢字雖然比部首檢字便利，比筆畫檢字有效率，但仍應再增加部首、筆畫等其他檢字方式，以求完備。四、漢字構形資料庫雖可透過網際網路免費下載，然而若要增加使用人次，應再開發中文繁簡體、英文、日文、韓文版或網路應用版。

上述問題的解決關鍵，其實就在網際網路。原先只想研發漢字構形資料庫的網路版以解決缺字、檢字及下載安裝的問題，然而若要建立形、音、義俱備的漢字知識庫，最好的方式還是將漢字構形資料庫的網路版擴充成文字學入口網站。文字學入口網站主要是利用漢字的形、音、義知識來檢字，並提供漢字字典網站的連結；如此一來，即可透過網際網路，集眾人之力來建立漢字知識庫並推廣漢字構形資料庫。規畫中的文字學入口網站應具有下列特色：一、多語的使用介面。二、多樣的檢字方式。三、可檢索不同字集。四、可解決缺字問題。五、提供字典網站連結。具有上述特色的文字學入口網站具有下列效益：一、推動字書上網。二、建立漢字知識庫。三、便利漢字知識的擷取。四、闡揚漢字文化。

文字學入口網站的規畫

莊德明 中央研究院資訊科學研究所研究助技師
derming@gate.sinica.edu.tw
鄧賢瑛 中央研究院資訊科學研究所研究助理
ying0419@iis.sinica.edu.tw

壹、漢字構形資料庫的研發與推廣

從殷商的甲骨文算起，漢字的使用已達 3,400 年之久；然而電腦及網路科技的蓬勃發展，只不過是近幾十年來的事。今天漢字在電腦中的應用，應該說只是滿足使用者基本的需求，還不完善，其最根本的問題為電腦中的漢字知識嚴重不足。中央研究院資訊科學研究所文獻處理實驗室自 1993 年起，開始進行漢字數位化的工作，我們的研究重心在於利用電腦建立漢字的知識庫，同時透過這些知識來提升電腦處理漢字的能力，而其中最主要的成果為建立了「漢字構形資料庫」。

漢字構形資料庫是一個記錄漢字形體知識的資料庫，這些知識包含古今文字的源流演變、字形結構及異體字表。自從 1998 年 8 月推出第一個正式版本後，其間為考量使用者的需求及漢字的特性，經歷了數次重大的改版更新，截至 2009 年 2 月份為止，所推出的版本已更新至 2.52 版；漢字構形資料庫的重要版本沿革，可參見表一。2008 年 5 月推出漢字構形資料庫 2.52 版，這個版本收錄古今漢字 116,790 個；其中楷書字形 62,942 個，小篆及重文 11,100 個，金文 21,413 個，楚系簡帛文字 19,138 個，甲骨文 2,197 個；另收《漢語大字典》異體字表 12,208 組。這個資料庫的主要特色如下¹：

1. 銜接古今文字以反映字形源流演變。
2. 收錄不同歷史時期的異體字表，以表達不同中文字在各個歷史層面的使用關係。
3. 記錄不同歷史時期的文字結構，以呈現中文字因義構形的特點。
4. 使用構字式及風格碼來解決古今文字的編碼問題。

我們已利用漢字構形資料庫開發漢字部件檢字系統，並用來解決缺字問題。經過這些年來的驗證，這個缺字問題的解決方案，基本上是可行的。目前「數位典藏與數位學習國家型科技計畫」至少已有 23 個資料庫採用漢字構形資料庫來解決缺字問題，其中規模最大的為中研院史語所漢籍電子文獻，漢籍電子文獻已上線古籍達三億五千八百萬字。在解決缺字問題方面，除了持續擴充漢字構形資料庫外，我們還進行下列工作：

1. 以公眾授權模式提供「中研院漢字部件檢字系統」的原始程式碼及相關資料，釋出給大眾使用。²

¹請參考〈漢字數位化的困境及因應：談如何建立漢字構形資料庫〉，莊德明，第五屆兩岸三院資訊交流與數位資源共享研討會，台北，2007 年 10 月

²請參考<http://cdp.sinica.edu.tw/cdphanzi/declare.htm>

2. 推動漢字構字標準，撰寫「中文字構形識別序列」標準草案，目前已函送標檢局審議。

表一、漢字構形資料庫的版本沿革

版本	年/月/日	內容簡介	字數	備註
1.0	1998/8/6	收錄五大碼字形 13,051 個。	13,051	1. 進行字形結構分析。 2. 研發部件檢字技術。
1.4	2002/05/17	1. 增收《漢語大字典》字頭，合計 54,711 個字形 2. 收錄《漢語大字典》異體字表 12,208 組。	54,711	1. 制定構字式以解決缺字問題。 2. 收錄異體字表。
2.1	2003/03/17	增收《說文解字詁林》的小篆及重文共 11,100 個。	65,811	收錄小篆，銜接古今漢字，並開始走向文字學資料庫。
2.2	2004/12/06	增收《金文編》中的 3,459 個金文及其楷定字。	73,779	收錄金文異構字。
2.3	2005/08/03	增收《楚系簡帛文字編》中的 2,627 個楚系簡帛文字及其楷定字。	77,268	收錄楚系文字異構字。
2.4	2006/8/2	增收《殷墟甲骨刻辭類纂》中的 1,970 個甲骨文。	81,720	收錄甲骨文異構字。
2.5	2007/8/9	1. 增收《金文編》中金文，總計 20,069 個金文。 2. 增收《楚系簡帛文字編》中的楚系文字，總計 16,801 個。 3. 增收《殷墟甲骨刻辭類纂》中的甲骨文，總計 2,197 個。	112,533	1. 大量增收金文及楚系簡帛文字的重文，合計 30,462 個。 2. 制定風格碼以解決古漢字的編碼問題。 3. 研發出處檢字技術。 4. 新增 Microsoft Word 自動貼圖功能。

雖然歷經十五年的開發，漢字構形資料庫不論在資料庫本身或推廣方面，仍有很大的改善空間：

1. 漢字有形、音、義三要素，漢字構形資料庫目前只著重在字形知識的整理，尚未涉及字音與字義；建立一個形、音、義俱備的漢字知識庫，仍是我們長遠的目標。
2. 漢字構形資料庫的缺字解決方案，採用單位仍侷限於中研院及「數位典藏與數位學習國家型科技計畫」，仍須再推廣。
3. 部件檢字雖然比部首檢字便利，比筆畫檢字有效率，但仍應再增加部首、筆畫等其他檢字方式，以求完備。
4. 漢字構形資料庫雖可透過網際網路免費下載³，然而由於檔案太大（約 84M），而且只能在繁體中文版下使用，以致下載人次增加有限。目前每天的下載人次約 10 次，總下載人次為 14,662 次（截至 2009 年 3 月 26 日

³ 漢字構形資料庫下載網址：<http://cdp.sinica.edu.tw/cdphanzi/>

止)。若要增加使用人次，漢字構形資料庫應再開發中文繁簡體、英文、日文、韓文版或網路應用版。

上述問題的解決關鍵，其實就在網際網路。原先只想研發漢字構形資料庫的網路版以解決缺字、檢字及下載安裝的問題，然而若要建立形、音、義俱備的漢字知識庫，最好的方式還是將漢字構形資料庫的網路版擴充成文字學入口網站。文字學入口網站主要是利用漢字的形、音、義知識來檢字，並提供漢字字典網站的連結；如此一來，即可透過網際網路，集眾人之力來建立漢字知識庫並推廣漢字構形資料庫。這樣一個入口網站不同於 Google、Yahoo 等用關鍵詞來找網頁的入口網站，而這樣的需求也非 Google、Yahoo 等入口網站所能達成。建立文字學入口網站的構想始於 2008 年初，接著我們開始搜尋網際網路的字典網站，瞭解這些網站的檢字方式及字典內容，同時也發現部分網站已具有文字學入口網站的概念。下一節先介紹幾個重要的網站，再來說明我們對文字學入口網站的規畫及實施。

貳、漢字字典網站簡介

一、《異體字字典》(網址：<http://dict.variants.moe.edu.tw/>)

《異體字字典》網站為教育部國語會所建置，介面採用繁體中文，收錄字數為 106,230 字 (2004 年 1 月版)，其中正字 29,892 字，異體字 76,338 字 (含待考之附錄字)。本字典所用的正字標準據教育部《常用字表》⁴、《次常用字表》⁵、《罕用字表》⁶等三正字表。若遇有三字表未收，而須獨立音義之文獻字形，則補收為《新正字表》。異體字收錄則參考《教育部異體字表》⁷。⁸

《異體字字典》採用部首及筆畫檢字，檢字後再列出該字的音讀及釋義。例如「𠵼」在作為正字時，讀音為ㄇㄟˇ，意思是「微嚐」；而當「𠵼」作為「吻」字的異體字時，則讀為ㄨㄣˇ，釋義同「吻」字。此外，在一旁的窗格中，還會列出該正字在《說文解字·大徐本》、《說文解字·段注本》或《龍龕手鑑》等參考資料中的說明。圖一為「𠵼」字的檢字界面，「𠵼」字為「口」部 5 畫。圖二為「𠵼」字可作為「吻」字的異體字。

在《異體字字典》網站中，字形的呈現方式是併用圖片與造字檔，因此使用者可能必需匯入該網站提供的造字檔，才能正確的瀏覽網站內容。同時因為含有大量的字形圖片，因此根據使用者的電腦等級及頻寬等條件不同，瀏覽的流暢度也會受到影響。

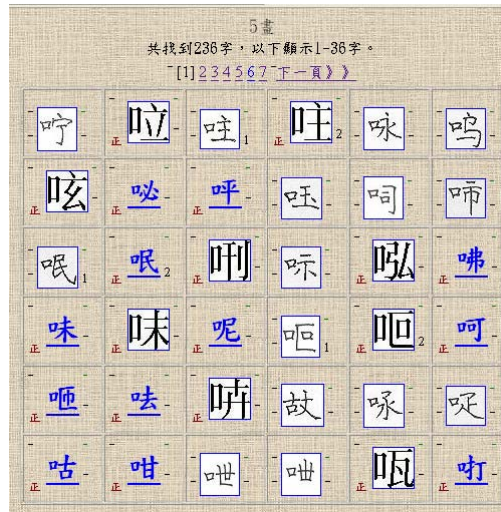
⁴ 此處指教育部於 1982 年 9 月交正中書局印行之《常用國字標準字體表》，共收常用字 4808 字。

⁵ 此處指教育部於民國 1982 年 10 月印行之《次常用字國字標準字體表》，共收次常用字 6334 字，並含 9 個單位詞，共計 6343 字。

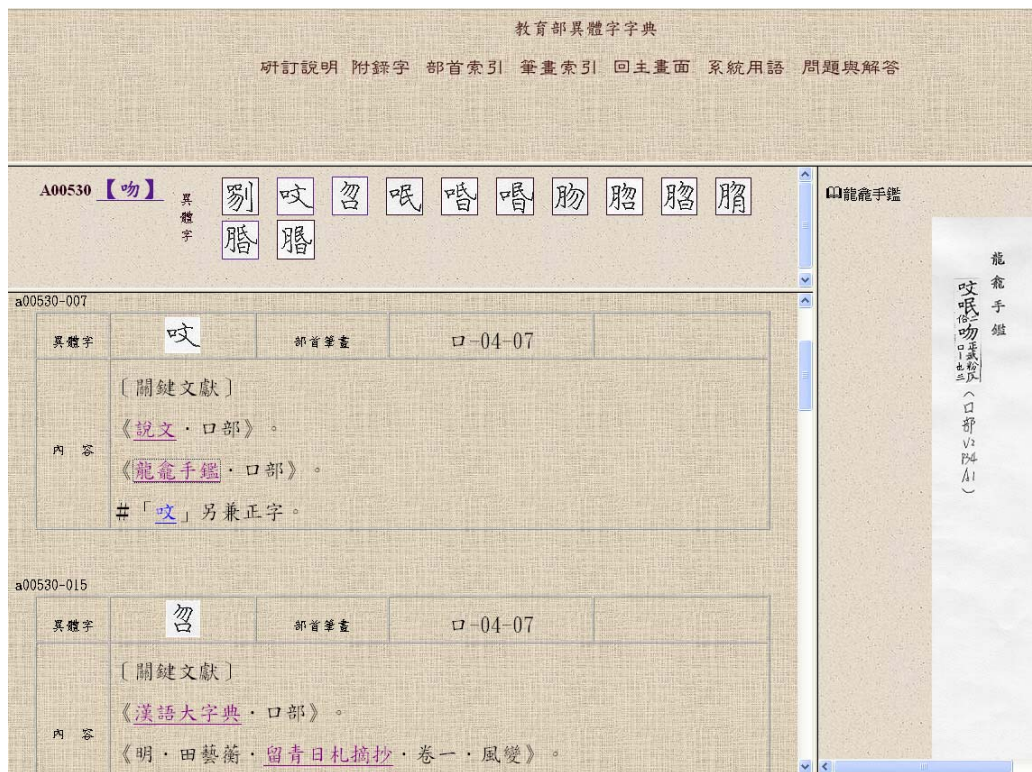
⁶ 此處指教育部於民國 1983 年 10 月印行之《罕用國字標準字體表》，共收罕用字 18388 字。

⁷ 此處指教育部於民國 1984 年 3 月印行之《異體國字字表》，共收異體 18588 字，補遺 22 字。

⁸ 以上資料參考教育部《異體字字典·編輯略例》(<http://140.111.1.40/bian/fbian.htm>)，以及《異體字字典·系統用語》(<http://140.111.1.40/ex.htm>)。



圖一、《異體字字典》「呬」字的檢字界面



圖二、《異體字字典》「呬」字為「呬」的異體字

二、CNS11643 中文標準交換碼全字庫（網址：<http://www.cns11643.gov.tw>）

CNS11643 中文標準交換碼全字庫是由行政院主計處電子處理資料中心所建置，介面採用繁體中文，5.0 版可供查詢的字共有 87,047 個中文字。查詢方式分為：筆畫、部首、注音符號、倉頡碼、拼音、CNS碼、部件⁹及筆順序¹⁰查詢八種選擇。以「槩」字為例，「槩」字的總筆畫為 19 畫；部首為「木」部；倉頡碼

⁹ 全字庫所依據的部件版本為CNS11643_2_94。

¹⁰ 全字庫所依據之筆順序版本為CNS11643_1_94

為「一日木」；漢語拼音為「zhu」；部件欄位可拆分爲「豕 艹 日 木」¹¹，上述方式都可以查詢到「彘」字。若使用者想要縮小查詢範圍，也可使用網站中的「複合查詢」，利用一種以上的條件進行查詢。圖三爲全字庫網站的查詢方式，圖四爲注音「ㄗㄨˋ」的查詢結果，圖五爲查詢到的「彘」字內容。

查詢到所需要的字形之後，便可將明體（點陣字形檔）與楷體（TrueType字形檔）的字形檔下載至個人電腦中使用。但使用者需安裝「全字庫軟體包」或「全字庫應用工具 4.0」，以便將字形檔下載回電腦後¹²，能夠利用上述軟體包中所提供的「字形轉入工具」，將字形匯入至個人電腦的「使用者造字區」之中，並且可由使用者自行指定碼位。¹³除了明體（點陣字形檔）與楷體（TrueType字形檔）以外，全字庫也提供歷代法帖的圖片檔以及字形的PNG檔，後者主要提供使用者在網頁即時顯示字形時所用。¹⁴

全字庫網站可說是提供了相當多元的查詢方式，使用者可使用不同的條件進行複合查詢，縮小查詢的範圍，節省查詢時間。然而在部件查詢中，使用者必需將字拆分至基礎部件，才能順利進行查詢；筆順序查詢對筆畫複雜的字，也不易作拆分；CNS 碼查詢，使用者則必需先知道該字所屬的字面以及字碼。這些查詢功能，都較難爲一般人所使用。另外，將字形下載回個人電腦，再匯入「使用者造字區」，此方法雖然解決了個人電腦中的字形顯示問題，然而卻並沒有徹底解決中文資訊交換的問題。



圖三、全字庫網站的查詢方式



圖四、全字庫注音「ㄗㄨˋ」的查詢結果

¹¹ 若使用者無法辨認出一個字中的所有部件，也可使用萬用碼(*)來代替不確定的部件。

¹² 在下載字形時，注意必需將該字存到C:\CMEX的子目錄內。

¹³ 查詢及匯入字形的詳細步驟，請參考http://www.cns11643.gov.tw/web/seek/seek_t.jsp。

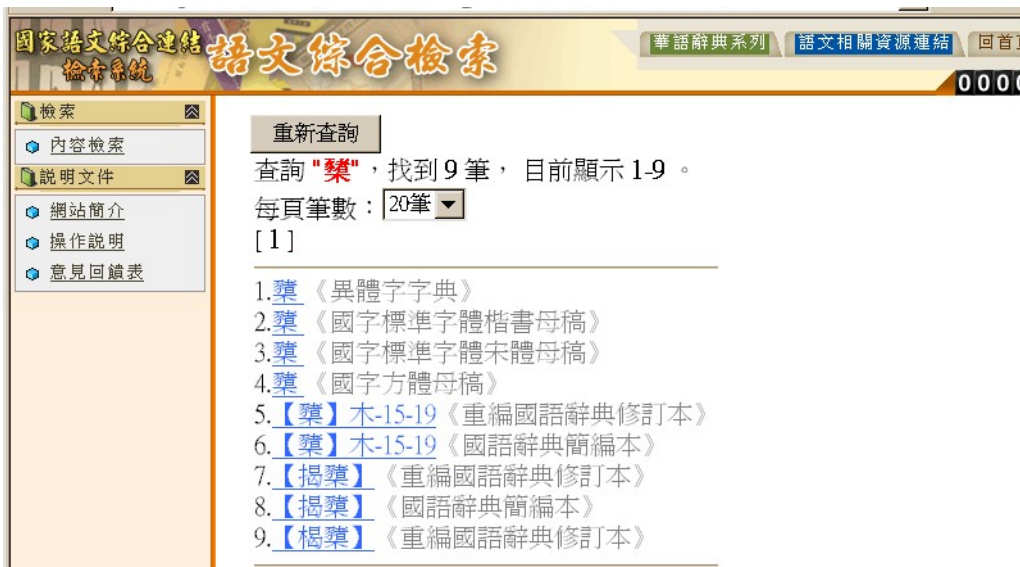
¹⁴ 全字庫的網頁即時顯示字形參數設定請參考<http://www.cns11643.gov.tw/web/share/share2.jsp>。

表二、國家語文綜合連結檢索系統所整合的資源

字典類	重編國語辭典修訂本
	國語辭典簡編本
	異體字字典
字形類	部首手冊
	國字標準字體楷書母稿
	國字標準字體宋體母稿
	國字標準字體教師手冊
	常用國字標準字體筆順手冊
	常用國字標準字體筆順學習網
	國字隸書母稿
	國字方體母稿
	國字標準字體研訂原則
	異體字表修訂版
字音類	國語注音符號手冊
	國語注音符號第二式
字義類	常用國字辨似
	常用量詞手冊
	新詞語料彙編 1(1996年7月~1997年12月)
	新詞語料彙編 2(1998年1月~1998年12月)
	新詞釋義彙編(一)
語料統計類	八十七年口語問卷調查報告書
	八十七年口語語料調查報告書
	八十七年常用語詞調查報告書
	八十六年常用語詞調查報告書
	八十五年常用語詞調查報告書
	八十四年常用語詞調查報告書
	國小學童常用字詞調查報告書
	大陸小學教科書字詞調查報告
	國語辭典簡編本編輯資料字詞頻統計報告
其他類	重訂標點符號手冊
	重編國語辭典(修訂本)編輯總報告書
	異體字字典編輯總報告書
	教育部成語典編輯總報告書



圖六、國家語文綜合連結檢索系統查詢「𪗇」字



圖七、國家語文綜合連結檢索系統查詢「𪗇」字的結果

四、漢字字典 (CCDICT) (網址：<http://chineselanguage.org/dictionaries/ccdict/>)

本網站的查詢介面有英文、繁體與簡體中文版三種。檢索方式分為：部首、拚音、廣東音、客語音、內碼、倉頡、四角號碼、英文、日文及韓文等方式。檢索結果可聯結至《漢英詞典》、《粵音音彙》¹⁶、《國語辭典》、《當代漢英詞典》等線上字詞典。圖八為漢字字典的查詢方式，圖九為查詢的結果及字典連結。

五、字典網 (網址：<http://zhongwen.com/zi.htm>)

本網站的檢字方式分為部首、內碼及英文，檢索的結果可連結至《粵音音彙》¹⁷、CEDICT¹⁸、教育部國語辭典¹⁹、CCDICT²⁰、Unihan Database²¹、CJK-English Dictionary、JDIC²²；此外，本網站還試圖囊括語源方面的檢索，可向外連結至 Tower of Babel²³、中央研究院瀚典全文檢索系統 1.3 版²⁴等檢索網站。圖十為字典網的查詢方式，圖十一為字典網「呵」的查詢結果及字典連結。

綜觀漢字字典及字典網這二個網站，除了一般常見的部首或內碼查詢以外，都加入了英文查詢，對其他非中文地區的使用者來說較為方便。漢字字典的檢索方式較為多元（有十種），但檢索結果所聯結的網站則較少。而字典網則相反，檢索方式雖然只有三種，但檢索結果聯結的網站則較多，而且範圍也較廣。

¹⁶ 此處指香港中文大學人文學科研究所所建置的黃錫凌《粵音音彙》電子版 (<http://humanum.arts.cuhk.edu.hk/Lexis/Canton/>)。

¹⁷ 同註 19。

¹⁸ Chinese-English Dictionary (<http://www.mandarin-tools.com/worddict.html>)。

¹⁹ 教育部國語辭典 (<http://dict.revised.moe.edu.tw/index.html>)。

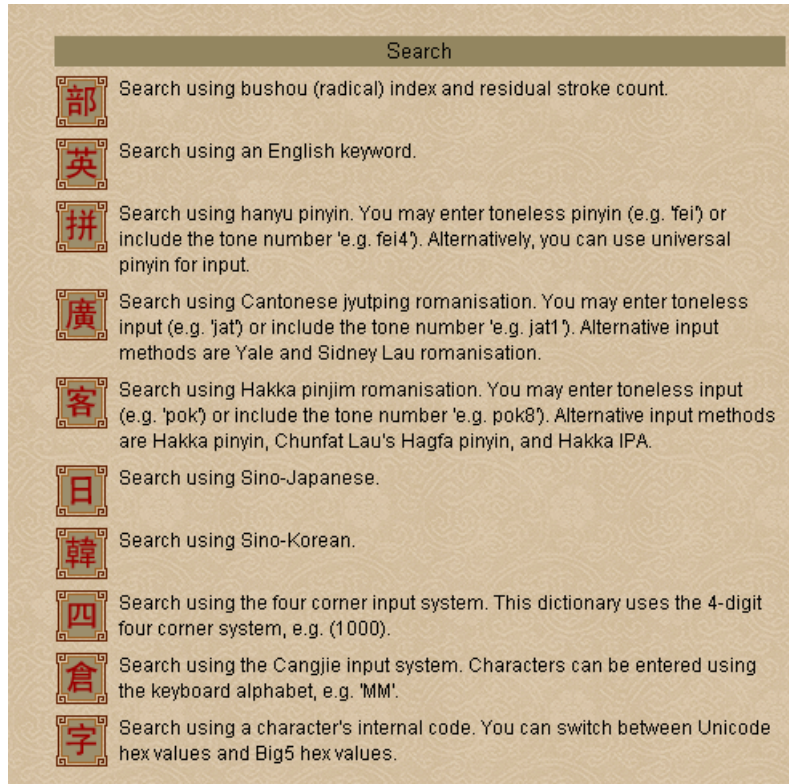
²⁰ 漢字字典 (CCDICT) (<http://chineselanguage.org/dictionaries/ccdict/>)

²¹ Unihan Database (<http://www.unicode.org/charts/unihan.html>)。

²² CJK-English Dictionary及JDIC (<http://www.csse.monash.edu.au/~jwb/wwwjdicinf.html>) 為日文的檢索網站。

²³ Tower of Babel (<http://www.omniglot.com/babel/chinese.htm>)。

²⁴ 中央研究院瀚典全文檢索系統 1.3 版 (<http://www.sinica.edu.tw/ftms-bin/ftmsw3>)



圖八、漢字字典的查詢方式

Home Forums Dictionaries Classroom Links Contact

Chineselanguage.org > Dictionaries > CCDICT > Bushou

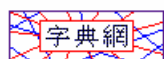
[正][簡]

Character	Dict. links	Bushou-strokes	Mandarin	English
同		口+3	tóng	[1] same; similar; equal; identical; common [2] [M] share; agree [3] together
同		口+3	tóng	
后		口+3	hòu hòu	[1] [n] queen, empress; sovereign [2] the god of the earth [3] after, behind [4] simplified form of 後
名		口+3	míng	[1] [n] name; rank; title; designation [2] position; honor; fame; renown; reputation [3] famous; noted; distinguished; renowned; valuable; precious; noble; rare, great [4] [n] name; [M] describe
名		口+3	míng	
各	廣州語 國語 當代	口+3	gè gé	[1] each; individually; every [2] all
各	廣州語 國語 當代	口+3	gè	
合	廣州語 國語 當代	口+3	hé xiá	[1] [M] combine; unite; join; gather; collect [2] [M] close; shut [3] [M] suit
合	廣州語 國語 當代	口+3	gè	
向	廣州語 國語 當代	口+3 卅+14	xiàng xiàng	[1] [M] turn; face [2] toward; [n] direction; trend [3] until now [4] a Chinese family name
吃	廣州語 國語 當代	口+3	chī jí jī	[1] [M] eat [2] [M] sustain; suffer; endure; bear [3] [WU] inhale; drink
吃	廣州語 國語 當代	口+3	qī	[M] stammer; stutter

圖九、漢字字典的查詢結果及字典連結



圖十、字典網的查詢方式



- From Unicode database:
Pronunciation: HE1
Definition: scold; laughing sound; yawn
- Calligraphy:
- [This character](#) in [Animated Chinese Characters](#). (simplified form)
- Dictionary entry:
- [This character](#) in [Cantonese Pronunciation Dictionary](#).
- [This character](#) in [CEDICT](#).
- [This character](#) in [Guoyu Cidian](#).
- [This character](#) in [CCDICT Chinese Character Dictionary](#).
- [This character](#) in [Unicode database](#).
- [This character](#) in [WWW CJK-English Dictionary](#). (Japanese)
- [This character](#) in [WWW JDIC](#). (Japanese)
- Etymology:
- [This character](#) in [Chinese Characters](#).

圖十一、字典網「呵」的查詢結果及連結

參、文字學入口網站的特色及效益

綜合上述，規畫中的文字學入口網站應具有下列特色：

1. 多語的使用介面。上述網站除了漢字字典為英文、繁體、簡體中文多語介面外，其餘皆只支援繁體中文或英文。作為一個入口網站，多語的介面是必須的，除了繁體、簡體中文、英文外，至少也須再兼顧其他漢字區的使用者，日文及韓文介面也須列入考量。
2. 多樣的檢字方式。漢字的形、音、義、字碼等屬性都可用來檢字。字形檢字包括部首、筆畫、部件等，字音檢字包括注音、拼音、聲韻等，字義檢字包括英文、日文、韓文等，字碼檢字包括 Big5、Unicode、CCII、CNS11643 等。
3. 可檢索不同字集。檢字技術好壞和字集大小息息相關，就如同輸入法一樣，字集一大，原先慣用的輸入法不見得能適用。例如字集變大了，同音、同筆畫的字數也隨著增加。入口網站應可讓使用者檢索不同的字集，這些字集包括常用字、簡化字，甚至甲骨文、金文、小篆等古漢字。
4. 可解決缺字問題。缺字指的是中文字碼裡沒有收錄的字，例如 Big5 收錄了 13,051 個字，13,051 之外的字即是缺字；現在常用的 Unicode 隨著不同版本字數雖然有所不同，少者 20,902 個字，多者可到 70,194 字，但是缺字問題仍然存在。無論使用者電腦中文字碼收錄字數的多寡，即使未安裝漢字字型，入口網站都應能正常顯示古今漢字。
5. 提供字典網站連結。目前字典網站彼此間的連結不夠，若要多瀏覽幾個網站，常需重複進行檢字，這對使用者而言並不方便。入口網站應在使用者找到所要的字後，提供現有的字典網站連結，以節省重複檢字的時間，並提供使用者更多的漢字參考資訊。

具有上述特色的文字學入口網站具有下列效益：

1. 推動字書上網。入口網站提供的檢字及缺字技術可降低字典上網的門檻，而入口網站提供的連結又可增加字典的能見度，如此不但可增加現有字典上網的意願，同時也有助於線上字典的編纂。
2. 建立漢字知識庫。古今字書的數位化工作，實非個別單位能獨立完成，如今正可透過網際網路集合眾人之力共同推動並予以整合，以建立完善的漢字知識庫。
3. 便利漢字知識的擷取。目前的入口網站主要透過關鍵字及關鍵詞來搜尋網頁，以漢字為主的網頁不易突顯；相形之下，文字學入口網站則只提供這些網頁的連結，因此更有利於使用者擷取到這些漢字知識。
4. 闡揚漢字文化。入口網站所提供的連結，除了字典網站外，也包括其他和漢字相關的文化、藝術網站，這些連結及網站可用來闡述及發揚漢字文化。

肆、進一步的工作

我們預計花兩年的時間來建立文字學入口網站，2009 年的工作規畫如下：

1. 持續擴充漢字構形資料庫以解決缺字問題。漢字構形資料庫將再增補楷書

- 缺字約四千個，甲骨文約五百個，總收錄字數將超過十二萬字。
2. 增補漢字屬性以增加檢字方式。輸入《漢語大字典》五萬多字的部首、筆畫數、古今聲韻等屬性，並研發部首、筆畫、古今聲韻檢字功能。
 3. 開發網路版應用程式以推廣漢字構形資料庫。開發漢字檢字網站，可讓使用者利用部件、部首、筆畫數、古今聲韻來檢索漢字，並協助罕用字的輸入。
 4. 撰寫技術報告《漢字構形資料庫的研發及應用》。
 5. 研發 Unicode 版漢字構形資料庫以支援中文繁簡體、英文、日文及韓文版的作業系統。

2010 年的工作規畫如下：

1. 建立 Unicode 版的漢字檢字網站，並執行文字學入口網站的檢字功能。
2. 將漢字構形資料庫收錄的金文及楚系文字轉換成《金文編》、《楚系簡帛文字編》的連結。
3. 加入教育部國語會《國語辭典》、《異體字字典》等字典的連結。

伍、結語

自 1993 年起開始進行漢字數位化的工作，至今已有 15 年的時間。這些年來，我們的重心都放在字形，基本上仍是在整理基本資料，雖然也發展了一些應用程式，主要還是用來解決缺字問題。解決缺字問題雖然有助於古籍文獻的數位化，但是應用的範圍仍有所侷限。對我們而言，建立文字學入口網站可說是一個新的開始，應用的範圍變廣，服務的對象變多，影響力自然也變大。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。最後，我們還是很高興有這樣一個機遇，能夠利用日新月異的數位化科技，透過網際網路集合眾人之力，來整合並傳播漢字的資訊，並進一步地闡揚漢字文化。

誌 謝

本論文得以順利發表，在此感謝國科會數位典藏與數位學習國家型科技計畫在經費上的支持，典藏系統支援及工具開發計畫（子計畫一）的計畫編號為 NSC 96-3113-H-001-011。