

# 《廣韻》字形的處理

2011年9月1日初稿

本文件旨在說明本資料庫如何處理《廣韻》中字形的呈現以及查詢功能。由於《廣韻》採用字形與現代常用字的落差、各地用字情形不一，及部分罕用字難以用輸入法打出等現象，本資料庫針對這些字形方面的問題，特設置了異體字檢索以及部件檢索機制。另外也對在不同作業系統下所產生的電腦缺字問題，進行個別處理。

漢字古今音資料庫網址：<http://xiaoxue.iis.sinica.edu.tw/ccr>

## 目錄

字頭.....	1
異體字.....	1
部件檢字.....	4
缺字的處理.....	6

## 字頭

漢字古今音資料庫所收錄的《廣韻》字頭共 25,528 個，其中字形重複共計 5,991 個，故實際收錄字形為 19,537 個。例如「揭」字在《廣韻》中對應 6 個字頭。

字號	字	聲調	韻目	字母	上字	下字	又音	又切
16510	揭	去	祭	溪	去	例		丘揭/其謁/居列/居揭/渠列
21714	揭	入	月	見	居	竭		渠列/丘揭/居列/其謁/去例
21718	揭	入	月	羣	其	謁		去例/丘揭/居列/居揭/渠列
22743	揭	入	薛	羣	渠	列	訃/竭	去例/其謁/居列
22782	揭	入	薛	溪	丘	竭		去例/居揭/其謁/渠列/居列
22885	揭	入	薛	見	居	列		居揭/去例/丘揭/其謁/渠列

## 異體字

由於《廣韻》採用字形與現代常用字的落差，加以各地用字情形不一，尤其海峽兩岸有繁簡字的用字差異，綜合以上時、空等變異因素，為了配合使用者的檢索需求，故提供異體字檢索功能。目前在《廣韻》的 25,528 個字頭中，增補異體字 4,501 個，扣除字形重複的 1,285 個外，實際支援檢索的異體字為 3,216 個。這 4,501 個異體字，其中 2,445

個為簡化字，1,460 個為《漢語大字典》<sup>1</sup>主體字，596 個為其他異體字。

但這些異體字目前並不會呈現在檢索結果的字頭欄位上。舉例來說，若以單字檢索簡化字「东风」，但在檢索結果中呈現的字形仍是《廣韻》字頭「東風」。

上古	中古	近代	現代	域外譯音
廣韻(中古)				
字號		字	东风	
部件		<input type="checkbox"/>	攝	
<input checked="" type="checkbox"/> 聲調		<input checked="" type="checkbox"/>	韻目	
<input checked="" type="checkbox"/> 字母		<input type="checkbox"/>	開合	
<input type="checkbox"/> 等第		<input type="checkbox"/>	清濁	
<input checked="" type="checkbox"/> 上字		<input checked="" type="checkbox"/>	下字	
<input checked="" type="checkbox"/> 又音		<input checked="" type="checkbox"/>	又切	
確定送出		重新設定		

在「字」的欄位中輸入簡化字「东风」

字號	字	聲調	韻目	字母	上字	下字	又音	又切
00001	東	平	東	端	德	紅		
00150	風	平	東	非	方	戎		方鳳
14736	風	去	送	非	方	鳳		方戎

在檢索結果中呈現的字形是《廣韻》字頭「東風」

又以常用字「窗」為例，實際上《廣韻》並未收錄此字，但透過異體字檢索機制，以單字檢索「窗」字，則能查出《廣韻》中收錄的「牕」、「窓」、「窻」等相關字頭，而使用者也能據此參考其相關的聲韻資料。

<sup>1</sup> 《漢語大字典》，徐中舒主編，遠東圖書公司，1991年9月。

上古 ▾ 中古 ▾ 近代 現代 ▾ 域外譯音

廣韻(中古)

字號	<input type="text"/>	字	<input type="text" value="窗"/>
部件	<input type="text"/>	<input type="checkbox"/> 攝	<input type="text"/>
<input checked="" type="checkbox"/> 聲調	<input type="text"/>	<input checked="" type="checkbox"/> 韻目	<input type="text"/>
<input checked="" type="checkbox"/> 字母	<input type="text"/>	<input type="checkbox"/> 開合	<input type="text"/>
<input type="checkbox"/> 等第	<input type="text"/>	<input type="checkbox"/> 清濁	<input type="text"/>
<input checked="" type="checkbox"/> 上字	<input type="text"/>	<input checked="" type="checkbox"/> 下字	<input type="text"/>
<input checked="" type="checkbox"/> 又音	<input type="text"/>	<input checked="" type="checkbox"/> 又切	<input type="text"/>

在「字」的欄位中  
輸入「窗」

字號	字	聲調	韻目	字母	上字	下字	又音	又切
00699	𪗇	平	江	初	楚	江		
00700	𪗈	平	江	初	楚	江		
00698	𪗉	平	江	初	楚	江		

可找出《廣韻》  
中與「窗」字有  
關的異體字字頭

透過異體字檢索機制，能夠減低因使用者輸入字形與《廣韻》收錄字形不同，以致無法檢索的機率，也能讓使用者以更直覺的方式檢索單字，不致發生連常用字都無法檢索的窘況。

## 部件檢字

部件是指字的一部分，當某些罕用字可能較不易以輸入法打出時，即可利用部件進行檢索。例如輸入部件「龍」，可檢索出「籠」、「籠」、「隴」、「礮」等這些包含部件「龍」的字。

上古 ▾ 中古 ▾ 近代 現代 ▾ 域外譯音

中古《廣韻》

字號  字

**部件 龍**  龕

聲調   韻目

字母   開合

等第   清濁

上字   下字

又音   又切

確定送出 重新設定

在部件的欄位中輸入「龍」

顯示項目 廣韻

共搜尋到 68 筆資料 / 7 頁

最初頁 1 2 3 4 5 6 7 下一頁 最末頁

字號	字	聲調	韻目	字母	上字	下字	又音	又切
00232	籠	平	東	來	盧	紅		力董/力鍾
00233	籠	平	東	來	盧	紅		
00234	礮	平	東	來	盧	紅		
00235	隴	平	東	來	盧	紅		
00236	礮	平	東	來	盧	紅		
00237	籠	平	東	來	盧	紅		力董
00238	籠	平	東	來	盧	紅		
00239	籠	平	東	來	盧	紅		呂江/所江
00240	礮	平	東	來	盧	紅		
00241	籠	平	東	來	盧	紅		

最初頁 1 2 3 4 5 6 7 下一頁 最末頁

檢索結果的總筆數有 68 筆

得出含有部件「龍」的字

使用者亦可輸入多個部件以縮小檢索範圍，例如在部件欄位中輸入「龍口」，檢索結果會送出包含部件「龍」及部件「口」的字，並且縮小檢索結果。

由於目前檢索結果的字頭並不列出異體字，因此部件檢字只能適用於字頭，不能檢索異體字。

上古 ▾ 中古 ▾ 近代 現代 ▾ 域外譯音

中古《廣韻》

字號  字

**部件 龍口**  繁

聲調   韻目

字母   開合

等第   清濁

上字   下字

又音   又切

在部件的欄位中輸入「龍口」

顯示項目 廣韻

共搜尋到 4 筆資料 / 1 頁

最初頁 1 最末頁

字號	字	聲調	韻目	字母	上字	下字	又音	又切
00236	𪛗	平	東	來	盧	紅		
00245	𪛘	平	東	來	盧	紅		
08170	𪛙	平	青	來	郎	丁		
09392	𪛚	平	覃	溪	口	含		

最初頁 1 最末頁

增加部件進行檢索後，檢索結果的總筆數減少成 4 筆

得出含有部件「龍口」的字

## 缺字的處理

此處所謂的「缺字」，是相對於系統預設支援的字集，即預設字集之外「不被支援的字形」，因此在不同的系統中，「缺字」的範圍也可能不同。例如 Windows 7 可支援的漢字個數為 70,207 個，而 Windows XP 只支援 20,914 個。《廣韻》實際收錄的 19,537 個字中，有 46 個為 Windows 7 的缺字，Windows XP 的缺字個數則為 2,720 個。

目前漢字古今音資料庫會依使用者的系統來處理缺字，網頁上的缺字則以圖片呈現。例如字號 20071 的「攸」字，Windows 7 可正常顯示，而 Windows XP 將以圖片呈現；又如字號 23439 的「輓」字，Windows 7 和 Windows XP 都以圖片呈現。

字號	字	聲調	韻目	字母	上字	下字	又音	又切
20071	攸	去	梵	溪	去	劍(劍)		
在 Windows XP 作業系統中，「攸」字為缺字，以圖片呈現								
20071	攸	去	梵	溪	去	劍(劍)		
在 Windows 7 作業系統中，「攸」字可正常呈現								
23439	輓	入	陌	羣	奇	逆		
在 Windows XP 和 Windows 7 作業系統中，「輓」字都為缺字，故均以圖片呈現								