

中文電腦缺字解決方案 漢字構形資料庫

莊德明 2004.10.11
中央研究院資訊科學研究所

Email:derming@gate.sinica.edu.tw

1

使用手冊目次(正文)

- ※ 第一章、系統安裝及移除
- ※ 第二章、系統功能簡介
- ※ 第三章、構字式的自動校正
- ※ 第四章、缺字字型
- ※ 第五章、實例輸入

2

使用手冊目次(附錄)

- ※ 附錄一、缺字的表達方式
- ※ 附錄二、構字符號的輸入
- ※ 附錄三、常用部件外字的輸入
- ※ 附錄四、安裝部件外字及輸入法

3

系統安裝及移除

4

系統需求

- ※ 個人電腦：IBM或100%相容的個人電腦
- ※ 作業系統：中文視窗95/98/ME/2000/XP
- ※ CPU：486以上
- ※ 主記憶體：32MB以上
- ※ 硬碟：100MB空間
- ※ 應用軟體：Microsoft Office 2000/XP中文版(建議)

5

安裝系統 6-1.a

1. 打開電腦的電源開關，啟動中文視窗作業系統。
2. 將「漢字構形資料庫」光碟片放入光碟機。
螢幕上會出現：

確定要安裝漢字構形資料庫？
3. 若要安裝，請按「確定」，否則按「取消」。

6

安裝系統 6-1.b

- ※ 若未出現上述的訊息，可選擇下面1或2任一方式來安裝系統：
1. 按一下螢幕左下方的「開始」，選擇「執行(R)」，在指令列中，請鍵入D:\setup，按「ENTER」（假設光碟機在D槽，若在E槽則鍵入E:\setup，餘依此類推）。
 2. 開啟「檔案總管」，在光碟機的根目錄內找到setup.exe，按兩下開始執行安裝程式。

7

安裝系統 6-2

1. 進入安裝程序後，畫面上會出現對話方塊：
安裝程式無法安裝使用中的系統檔案，也無法更新使用中的的共用檔案。建議您在安裝前，關閉所有的應用程式
2. 關閉其他的應用程式，然後按「確定」。若不想安裝，則按「結束安裝」。

8

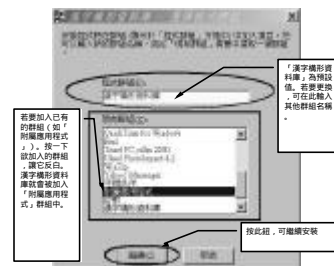
安裝系統 6-3

- ※ 按確定之後，出現以下畫面，按一下最左上角的按鈕，開始安裝。



9

安裝步驟 6-4



10

安裝步驟 6-5

1. 按「繼續」後，安裝程式開始抄錄檔案，抄錄完成時，畫面出現下列訊息：

安裝完成！

2. 按下「確定」，即可完成安裝程序。

11

安裝步驟 6-6

- ※ 安裝成功後，出現以下訊息時，請再重新啟動電腦。

漢字構形資料庫必須在電腦重新啟動後才能使用，現在是否要重新啟動電腦？（假使缺字字型在開機後無法正常顯示，請先改用安全模式開機，再以正常模式重新啟動，即可恢復正常。）

12

移除系統

1. 按「開始」、「設定」、「控制台」
2. 按兩下「新增/移除程式」
3. 選擇「漢字構形資料庫」，按「新增/移除」
4. 按「確定」按鈕執行移除動作
5. 移除程式會再詢問您是否確定要移除，請按「確定」
6. 移除完畢後請按「確定」按鈕

13

安裝須知 1：先移除舊系統

- ※ 若曾經安裝過本系統，即先移除先前的版本後再重新安裝。

14

安裝須知 2.1：重新開機

1. 安裝過程中如果顯示以下訊息：

因為某些系統檔並非最新版，導致無法繼續安裝。假如您希望程式自動更新這些檔案，請選「確定」；您必須重新開機後，才能再執行安裝程式。假如您不希望變動這些檔案，請選擇「取消」。

2. 選擇「確定」之後，繼續出現以下訊息：

15

安裝須知 2.2：重新開機

要現在啟動視窗系統嗎？如果您選擇「否」，則必須等到下一次啟動後，才能再執行安裝程式。

3. 此時請選擇「確定」，讓電腦重新開機。然後再重新執行一次光碟片上的 setup.exe。

16

安裝須知 3：版本衝突

- ※ 若出現以下的「版本衝突」訊息，請保留目前擁有的檔案。

現在正在複製的檔案比您系統目前擁有的檔案要舊，建議您保留目前擁有的檔案。要保留此檔案嗎？

17

安裝須知 4：檔案複製錯誤

- ※ 若出現以下的「檔案複製錯誤」訊息，請選取「忽略」，並繼續安裝。

複製 xxx 檔案時，發生存取錯誤

- ※ 原因：xxx 檔案正在使用中，以致無法更新。

18

安裝須知 5：缺字字型無法顯示

※ Win95/98由於採用字型快取技術(存在於Windows目錄中的ttfcache檔)，一旦缺字字型無法正常顯示，請依下列方式解決：

1. 先改用安全模式開機。
2. 再以正常模式重新啟動。

19

安裝須知 6：外字檔及輸入法



20

安裝須知 7.1：Win2000/OfficeXP

※ Microsoft Windows 2000的電腦中一旦安裝Microsoft Office XP後，假使再外掛外字輸入法，會導致無法開機或關機。請依下列步驟安裝：(Win2000 / Office2000、WinXP / Office2000、WinXP / OfficeXP無類似問題)

1. 參考下圖取消外字輸入法的安裝後，再重新啟動電腦。

21

安裝須知 7.2：Win2000/OfficeXP



22

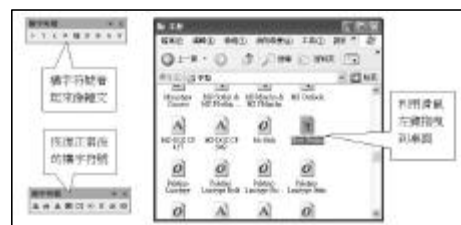
安裝須知 8.1：構字符號像韓文

※ Office XP安裝的韓文字型New Gulim 會和構字符號衝突，請先移除該字型，構字符號即恢復正常。

1. 開啟「控制台」的「字型」資料夾，再參考下圖移除New Gulim 字型。

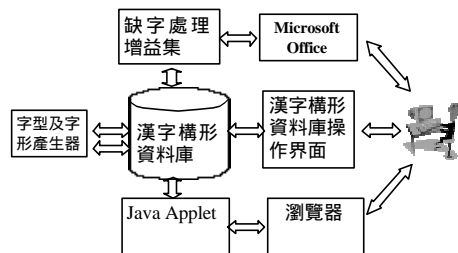
23

安裝須知 8.2：移除韓文字型



24

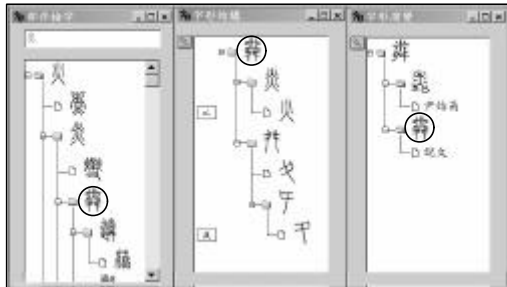
✳ 上述的安裝問題，可詳見常見問題集。



甲骨文	金文	楚簡	小篆	楷書
字集	↔ 字集	↔ 字集	↔ 字集	↔ 字集
部件集	部件集	部件集	部件集	部件集
字根集	字根集	字根集	字根集	字根集
異體字表	異體字表	異體字表	異體字表	異體字表
電腦字型	電腦字型	電腦字型	電腦字型	電腦字型

形體關係	形義或使用關係	直接部件
𤇓 𤇓 尹姑商	人死後身旁出現的焚火	𤇓
𤇓 說文	兵死及牛馬之血為𤇓。 𤇓，鬼火也，从炎𤇓。	炎、𤇓
𤇓	磷火。後作「磷」。	米、𤇓
磷	一種非金屬元素	石、𤇓

小篆的部件檢字及解形



31

金文構形資料庫

- 收錄字形以《金文編》第四版為主
- 《金文編》正文包含字頭2,420個、字形21,737個，其中異構字形約3,500個。附錄另有2,483個尚未辨識的字形。
- 採用中研院史語所金文工作室的金文構形分析來登錄字形結構
- 金文的部件為804個，字根為469個。
- 提供中研院金文字型
- 目前已建置異構字形3,454個

32

金文的部件檢字及異構字



33

金文異寫字(得)



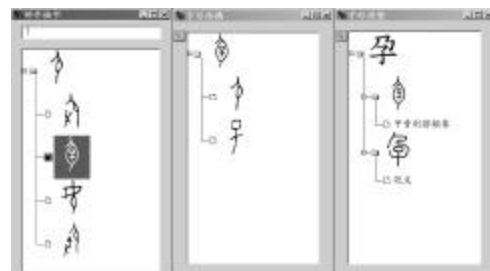
34

甲骨文構形資料庫

- 收錄字形以《殷墟甲骨刻辭類纂》為主。
- 《殷墟甲骨刻辭類纂》收錄字形約4,500個，目前只建置177個。

35

甲骨文的部件檢字及字形演變



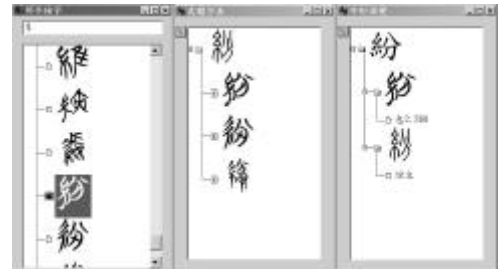
36

楚系文字構形資料庫

- ※ 收錄字形以《楚系簡帛文字編》為主。
- ※ 《楚系簡帛文字編》收錄字頭2,228個，字形19,250個。
- ※ 目前只建置異構字形170個。

37

楚系文字的部件檢字及異構字



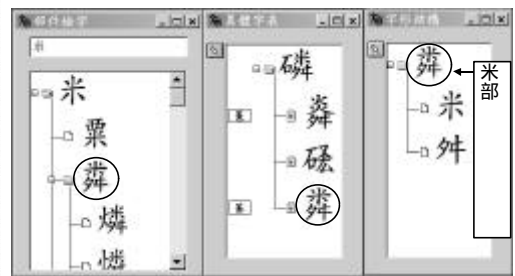
38

楷書構形資料庫

- ※ 收錄字形以《漢語大字典》為主，另外增收金文等古文楷化字，目前總字數為59,076個。
- ※ 參考《康熙字典》的部首分類來拆分字形。
- ※ Big5及簡化字的部件個數為2,161個、字根為529個，可涵蓋約57,000個字。
- ※ 提供Big5楷體、明體缺字各46,018個

39

楷書的部件檢字、異體字及解形



40

異體字表

- ※ 已收錄《漢語大字典》的異體字表，共12,208組，涵蓋36,309個字形。
- ※ 楷書的異體字表事實上是一個反映各個時期漢字的異體字表的聯集。
- ※ 異體字表提供古今字、簡繁字對照
- ※ 目前異體字表共12,442組，涵蓋38,048個字形。

41

異體字表的古今字、簡繁字



42

古文字型及缺字字型



備註：每套外字集最多含6,217個外字

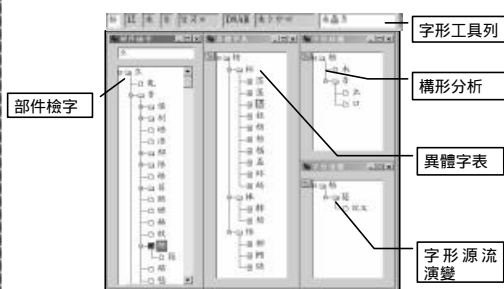
43

缺字公用程式



44

漢字構形資料庫(操作界面)



45

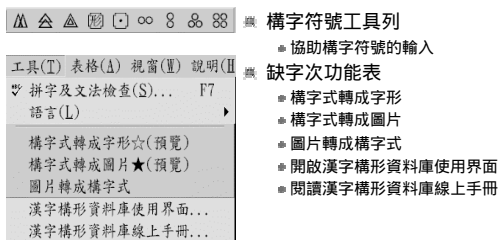
使用Microsoft Office 處理缺字

.....

.....

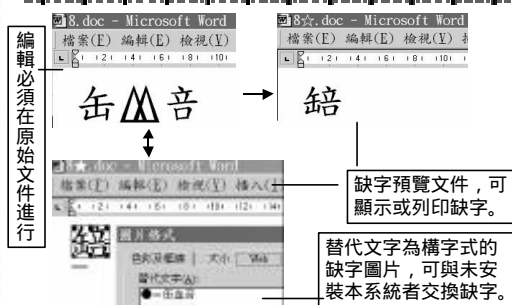
46

Microsoft Office缺字增益集



47

構字式、字形及缺字圖片的轉換



48

原始文件及預覽文件

- ✎ 使用者輸入構字式的文件，稱為「原始文件」。
- ✎ 執行缺字轉換後，可將構字式轉成字形，並存於新增的文件中，這個新增的文件稱為「缺字預覽文件」。
- ✎ 缺字預覽文件名稱為原始文件名稱再附加「」或「」。
- ✎ 由於構字式可當交換碼，所以原始文件適合交換及修改。
- ✎ 缺字預覽文件只供閱讀及列印，修改時必須回到原始文件進行。

49

構字式、字型及缺字圖片的比較

	純文字	圖形環境	字形顯示	字形列印	編輯	額外安裝
構字式	可用	可用	×	×	建議	部件外字
字形(型)	×	可用	優	優	不建議	字型
缺字圖片	×	視軟體版本而定	優	視解析度而定	不建議	無

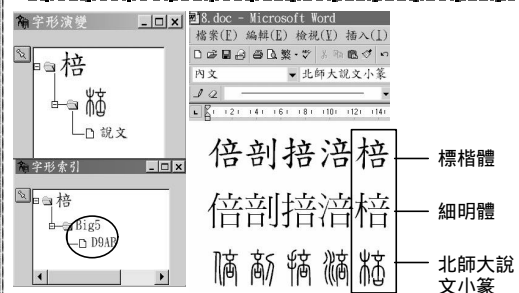
50

Office 2000/XP 的缺字處理比較

	Word	Excel	Power Point	Access	Front Page
構字式	支援	支援	支援	支援	支援
字形(型)	支援	支援	支援	×	支援
缺字圖片	支援	×	×	×	支援

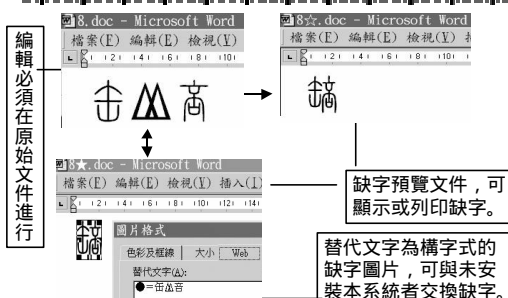
51

使用小篆字型 1/2



52

使用小篆字型 2/2



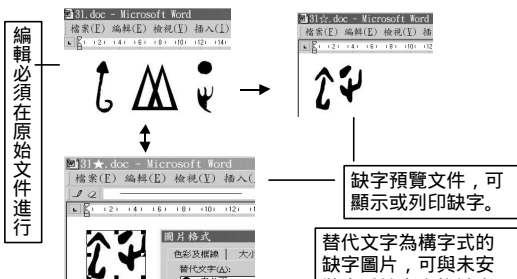
53

使用金文字型 1/2



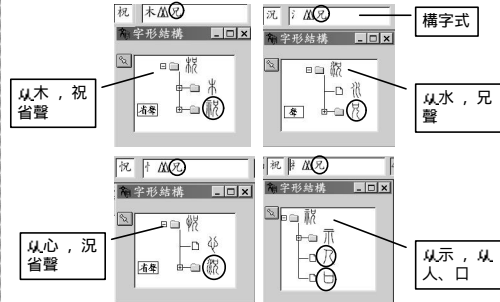
54

使用金文字型 2/2



55

構字式反映楷體字形結構



56

在Word處理缺字(實例操作)

- ✧ 構字式轉成字形(手冊第12頁、例2.1)
- ✧ 構字式轉成圖片(手冊第13頁、例2.2)
- ✧ 檢視圖片中的構字式(手冊第14頁、例2.3)
- ✧ 圖片轉成構字式(手冊第16頁、例2.4)
- ✧ 解除文件保護(手冊第17頁、例2.5)
- ✧ 編輯文件內容(手冊第19頁、例2.6)
- ✧ 構字式和一般文字(手冊第20頁、例2.7)
- ✧ 構字輸入法(手冊第20頁、例2.8)

57

在Excel處理缺字(實例操作)

- ✧ 構字式轉成字形(手冊第21頁、例2.9)
- ✧ 在Excel中，只能將構字式轉成字形，無法轉成圖片檔，也無法將圖片轉成構字式。

58

在PowerPoint處理缺字(實例操作)

- ✧ 構字式轉成字形(手冊第22頁、例2.10)
- ✧ 在PowerPoint中，只能將構字式轉成字形，無法轉成圖片檔，也無法將圖片轉成構字式。

59

在FrontPage處理缺字(實例操作)

- ✧ 構字式轉成圖片(手冊第22頁、例2.10)
- ✧ 如果不先將構字式轉成圖片，在網頁上只能看到構字式，無法看到字形。

60

簡介

61

字形源流演變中的小篆

字頭

解形

釋義

《漢語大字典》的「梔」字

62

字形源流演變中的小篆 說文或體

字頭

解形

釋義

《漢語大字典》的「耆」字

63

部件檢字

字形工具列

構形分析

異體字表

字形源流

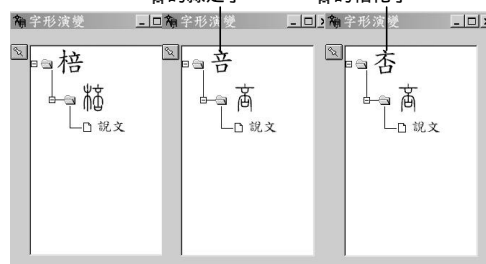
64

查詢「梔」時可輸入：梔、木𣎵音、木音、木立口



65

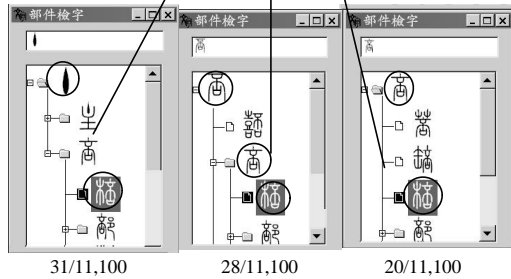
商	商
的隸定字	的楷化字



66

部件檢字(小篆) 2/3

依說文解字部首排序



31/11,100

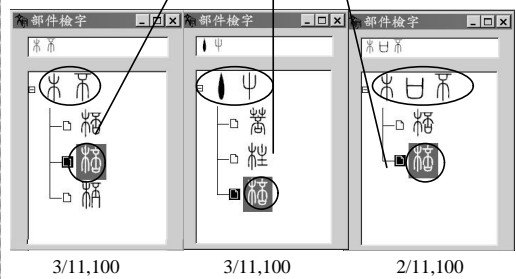
28/11,100

20/11,100

73

部件檢字(小篆) 3/3

依說文解字部首排序



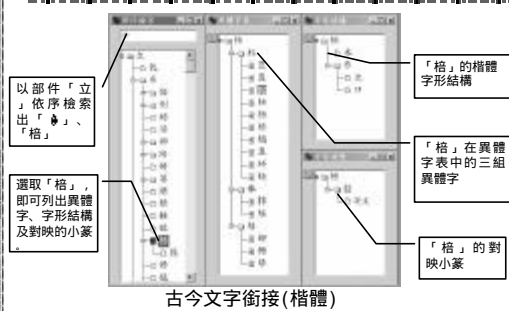
3/11,100

3/11,100

2/11,100

74

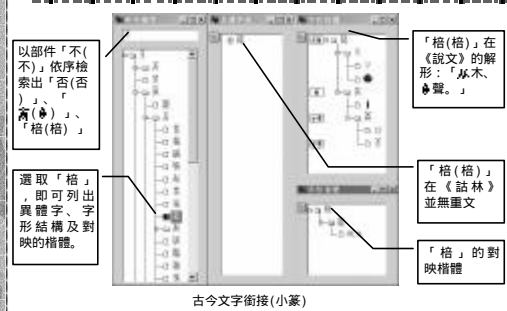
古今文字銜接機制(一)



古今文字銜接(楷體)

75

古今文字銜接機制(二)



古今文字銜接(小篆)

76

開啟操作界面

- ✱ 按一下Word、Excel、PowerPoint及FrontPage中的功能表「工具」，選「漢字構形資料庫使用界面」，就可以開啟漢字構形資料庫使用者界面的視窗。
- ✱ 按一下Windows工作列上的「開始」，再按「程式集」，再按「缺字公用程式」，選「漢字構形資料庫」，就可以開啟漢字構形資料庫使用者界面的視窗

77

漢字構形資料庫 實例操作

- ✱ 部件檢字(手冊第27-29頁、例2.12-2.14)
- ✱ 複製構字式(手冊第30頁、例2.15)
- ✱ 查詢異體字(手冊第31頁、例2.16)
- ✱ 查詢字形源流演變(手冊第32頁、例2.17)
- ✱ 查詢字形結構(手冊第34頁、例2.18-2.19)

78

網際網路的缺字處理

79

網頁的缺字表達

- ✱ 採用缺字圖片表達網頁的缺字。
- ✱ 瀏覽者的電腦不需額外安裝軟體或字型，即可顯示缺字。
- ✱ 瀏覽者的電腦若安裝本系統，即可將圖片轉成構字式，以作進一步的運用。
- ✱ 編製網頁時，不受字集的限制，可自由的表達缺字並與瀏覽者交換。

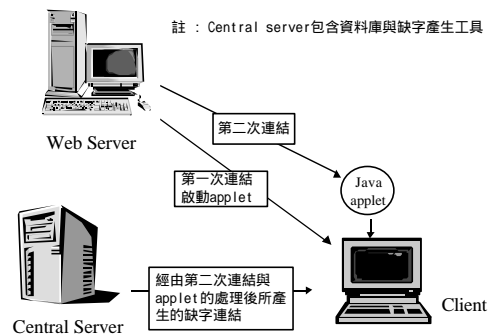
80

網頁的缺字表達(實例操作)

- ✱ FrontPage中的缺字圖片 (第36頁、例2.20)
- ✱ 缺字圖片的標籤(第37頁、例2.21)

81

使用Java Applet轉換網頁中的構字式

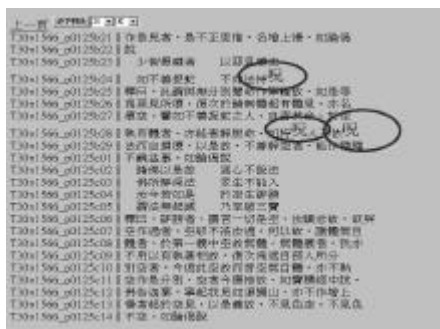


使用說明

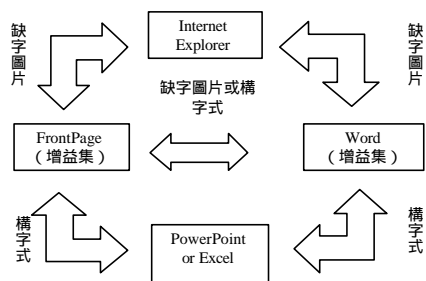


藍色所圈選之處為缺字所在，紅色圈選之處為轉換缺字的按鈕，並且可選擇所產生缺字的大小與顏色。

轉換後的結果，若要還原請重新載入即可



使用Microsoft Office處理網頁缺字



85

構字式的自動校正

.....

86

自動校正(實例操作)

- ※ 一般文字的構字式
- ※ 同一字形的不同構字式(手冊第46-47頁、例3.1-3.2)
- ※ 字形相似的構字式(手冊第48頁、例3.3)
- ※ 包含異寫部件的構字式(手冊第49頁、例3.4)

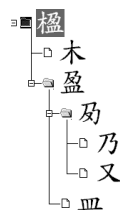
87

缺字的表達方式

.....

88

部件



- ※ 漢字的構形單位稱為部件，如：
 - 「木」和「盈」為「楹」的部件
 - 「乃」和「皿」為「盈」的部件
 - 「丩」和「乚」為「乃」的部件
- ※ 部件是有層次的，如左圖。

89

構字符號

- ※ 構字符號可表達部件連接的方式，也是系統辨識一般文字和缺字的依據。
- ※ 構字符號有三類，共計十三個：
 - ◆ 連接符號：𠂔、𠂕、𠂖
 - ◆ 起止符號：𠂗、𠂘
 - ◆ 方便符號：𠂙、𠂚、𠂛、𠂜、𠂝、𠂞、𠂟、𠂠、𠂡、𠂢、𠂣

90

連接符號

- **橫連(厶)：**
當部件的連接順序是由左至右，如：桃=木厶兆。
- **直連(厶)：**
當部件的連接順序是由上至下，如：李=木厶子。
- **包含(凵)：**
當部件的連接順序是由外至內，如：困=凵厶木。

91

起止符號

- 在不使用連接符號的情況下，只用部件來表達缺字，必須在部件前後加上起始符號(◻)和終止符號(◻)包夾，如：𪗇=◻𪗇◻ 𪗇域哲◻。

92

方便符號

- 魚 ∞ : 二個相同部件橫連，如：𩺰=∞魚。
 魚 8 : 二個相同部件直連，如：𩺠=8魚。
 魚 ∞∞ : 三個相同部件橫連。
 魚 88 : 三個相同部件直連。
 魚 𩺠𩺠 : 三個相同部件呈三角狀排列，如：𩺠=𩺠魚。
 魚 ∞∞∞ : 四個相同部件橫連。
 魚 888 : 四個相同部件直連。
 魚 𩺠𩺠𩺠 : 四個相同部件呈四角狀排列，如：𩺠=𩺠魚。




93

構字式

- 構字式可由部件連接方式區分成橫連、直連及包含三種。
- 為方便使用者輸入，除了以上三種構字式外，還有部件序及方便符號可以使用。



94

橫連、直連、包含

- 橫連：**
一個字形中的部件是用橫向的方式連接，如：
桃=木兆。
- 直連：**
一個字形中的部件是用直向的方式連接，如：
李=木子
- 包含：**
一個字形中的部件被另一個部件包含，如：
困=口木。

95

部件序(一)

- 採用部件連接的方式輸入構字式時，如果遇到不會輸入或電腦字集沒有的部件，可以將該部件拆分成更小的部件來輸入。此時只需按照部件的書寫順序輸入，前後以「」和「」包夾即可。如：

趁=形 域哲+

96

構字符號的輸入(倉頡及注音) 2/2

符號	倉頡	注音	備註
∞	竹竹(HH)	□	H:horizontal (水平的)
∞∞	竹竹竹(HHH)		
∞∞∞	竹竹竹竹(HHHH)		
8	難難(XX)	□	很難輸入
88	難難難(XXX)		
888	難難難(XXX)		
8888	難難難難(XXXX)		
88888	難難難難難(XXXXX)		

103

部件外字的輸入(倉頡及注音)

- 部件外字的倉頡碼，取碼方式為：「難+首尾碼」。「難」表「很難輸入」，也可和電腦用字做一區隔。如「𠂇」的倉頡碼為「難口一」。若只有一碼可取，則只要取一碼即可，如「イ」的倉頡碼為「難人」。
- 部件外字的注音碼同一般電腦用字，如「𠂇」的注音碼為「ㄇㄨˊ」、「イ」的注音碼為「ㄧˊ」。

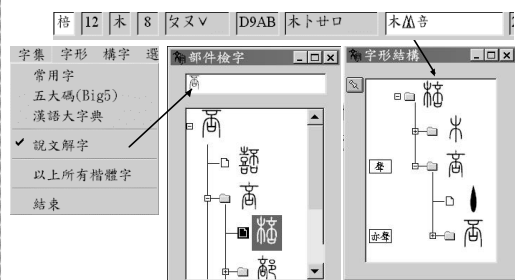
104

漢字構形資料庫操作界面

進階

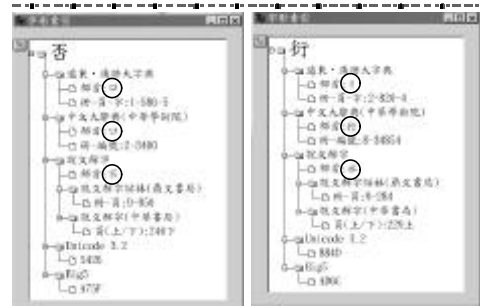
105

字集功能表



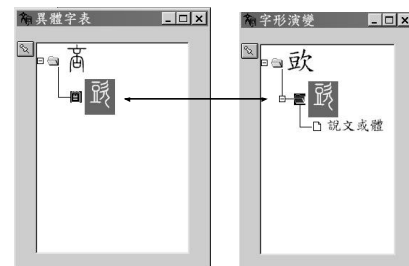
106

字形索引



107

小篆重文



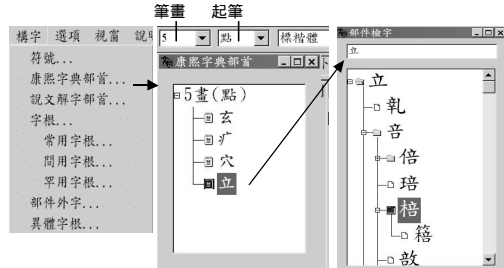
108

漢字時空座標



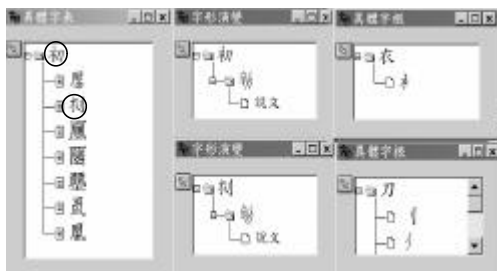
109

構字部件



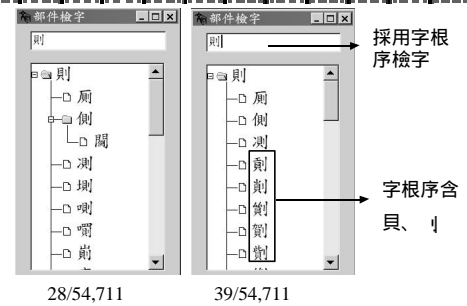
110

異體字根



111

部件檢字只列字形不列部件

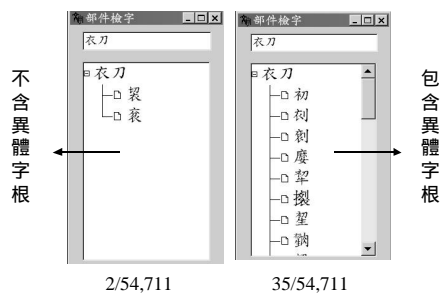


28/54,711

39/54,711

112

部件檢字包含異體字根

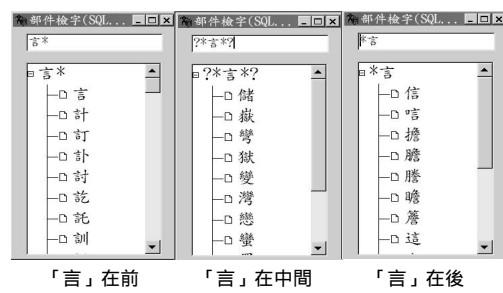


2/54,711

35/54,711

113

部件檢字採用SQL Like語法



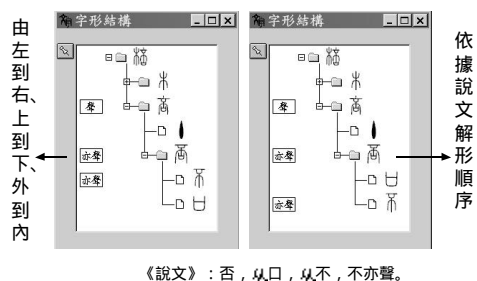
「言」在前

「言」在中間

「言」在後

114

依據說文解形順序列出小篆部件



115

開啟預設視窗(基本)



116

開啟預設視窗(進階)



117

感謝您的聆聽
和耐心

您的批評和指教
正是我們改進的契機

118