

智慧型文字編碼 漢字缺字的解決方案

莊德明 2003.04.28

中央研究院資訊科學研究所

Email:derming@gate.sinica.edu.tw

1

大綱

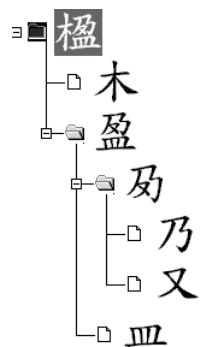
- ✧ 缺字的表達方式
- ✧ 系統安裝及移除
- ✧ 系統功能簡介
 - 使用Microsoft Office處理缺字
 - 漢字構形資料庫操作界面
 - 網際網路的缺字處理
- ✧ 構字式的自動校正

2

缺字的表達方式

3

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
- ◆ 「乃」和「又」為「乃」的部件

※ 部件是有層次的，如左圖。

4

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：△、△、△

● 起止符號：◻、◻

● 方便符號：○○、○○○、○○○○、○、○○、○○○、○○○、
○○

5

連接符號

✱ 橫連(△)：

當部件的連接順序是由左至右，如：桃=木△兆。

✱ 直連(△)：

當部件的連接順序是由上至下，如：李=木△子。

✱ 包含(△)：

當部件的連接順序是由外至內，如：困=口△木。

6

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：儻 = ◻儻◻ 西域哲◻。

7

方便符號

✱ ○○：二個相同部件橫連，如：鱻 = ○○魚。

✱ 〇：二個相同部件直連，如：鱻 = 〇魚。

✱ ∞：三個相同部件橫連。

✱ 〇：三個相同部件直連。

✱ 〇〇：三個相同部件呈三角狀排列，如：鱻 = 〇〇魚。

✱ ∞∞：四個相同部件橫連。

✱ 〇〇：四個相同部件直連。

✱ 〇〇〇〇：四個相同部件呈四角狀排列，如：鱻 = 〇〇〇〇魚。

8

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。



9

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 Δ 兆。
- ✱ 直連：
一個字形中的部件是用直向的方式連接，如：
李=木 Δ 子
- ✱ 包含：
一個字形中的部件被另一個部件包含，如：
困=口 Δ 木。

10

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「形」和「」包夾即可。如：

儷=形 西域哲

11

部件序(二)

- ✱ 由於某些字形的部件書寫順序並不明確，為了兼顧使用上的方便，部件的順序不影響缺字的辨識。如：

儷=形西域哲人

12

方便符號(一)

- ※ 部分字形會重覆使用到某一部件，如「鯪」的構字式為「魚△魚」，可用方便符號「○○魚」來表示，較為精簡。
- ※ 方便符號的表達要列在部件的前面，如「○○魚」，不能以「魚○○」來表示。
- ※ 方便符號可以使用在前述四種構字式中，如後：

13

方便符號(二)

類別	例字	構字式
橫連	俎	㇀人△且
直連	蠶	○○死△日△○○虫
包含	孱	尸△品子
部件序	歡	形○○又酉欠□

14

構字式注意事項

- ✱ 橫連、直連或包含的構字符號在構字式中只能使用一種。若要使用一種以上的構字符號，必須改用部件序，如
儷 ≠ 亻 厶 西 厶 域 厶 哲，儷 = 形 亻 西 域 哲
- ✱ 同一種橫連、直連或包含的構字符號在構字中可重覆使用，如
嶺 = 山 厶 豕 厶 〇 土

15

構字符號 及部件外字的輸入

16

構字符號的輸入(倉頡及注音) 1/2

符號	倉頡	注音	備註
⊖	重日(ZA)	ㄅ	形狀像倒下來的B
⊗	重月(ZB)		形狀像A
⊕	重金(ZC)		C:contain(包含)
⊠	重竹(ZH)	ㄉ	「形」的倉頡尾碼為「竹」
⊡	重戈(ZI)		點的倉頡尾碼為「戈」

17

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)	冂	H:horizontal (水平的)
○○○	竹竹竹(HHH)		
○○○○	竹竹竹竹(HHHH)		
⊙	難難(XX)		很難輸入
⊙⊙	難難難(XXX)		
⊙⊙⊙	難難難(XXX)		
⊙⊙⊙⊙	難難難難(XXXX)		
⊙⊙⊙⊙	難難難難(XXXX)		

18

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「𠂔」的倉頡碼為「難口一」。若只有一碼可取，則只要取一碼即可，如「亻」的倉頡碼為「難人」。
- ✱ 部件外字的注音碼同一般電腦用字，如「𠂔」的注音碼為「ㄇㄨㄛˊ」、「亻」的注音碼為「ㄇㄨˊ」。

19

系統安裝及移除

20

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)

21

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」。

22

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。

23

安裝系統 6-2

1. 進入安裝程序後，畫面上會出現對話方塊：

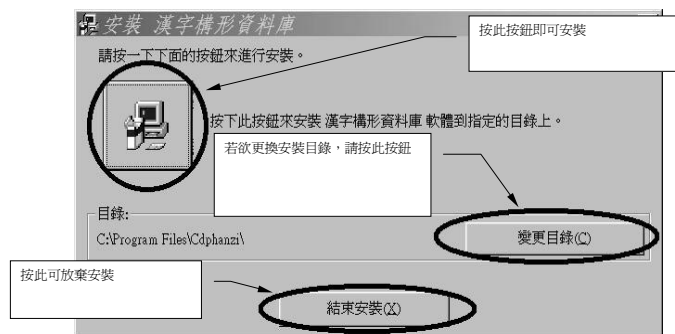
安裝程式無法安裝使用中的系統檔案，也無法更新使用中的的共用檔案。建議您在安裝前，關閉所有的應用程式

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」。

24

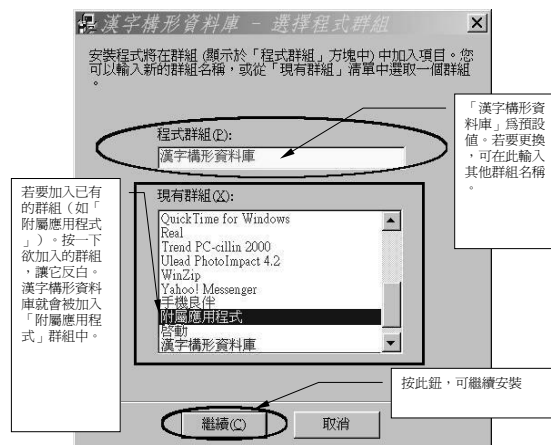
安裝系統 6-3

✳ 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。



25

安裝步驟 6-4



26

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。

27

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。）

28

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

29

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。

30

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：

31

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。

32

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？

33

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。

34

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。

35

安裝須知 6：外字檔及輸入法



36

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. 當漢字構形資料庫安裝完畢後，當安裝程式詢問是否要重新啟動電腦，請按「取消。」
2. 參考下圖取消外字輸入法的安裝後，再重新啟動電腦。

37

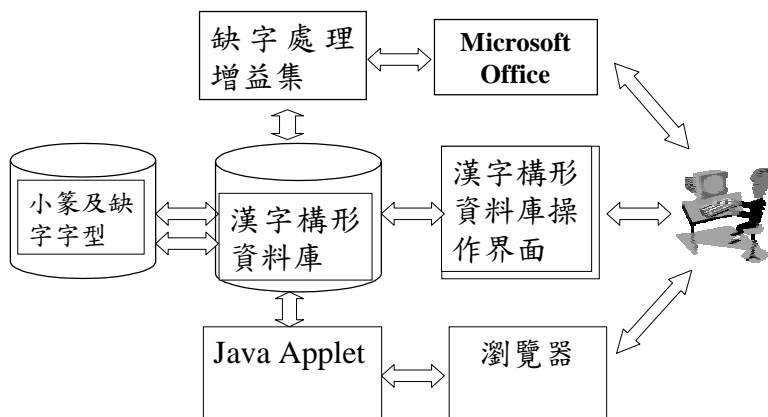
安裝須知 7.2：Win2000/OfficeXP



38

系統功能簡介

系統架構



漢字構形資料庫 1/2

- ※ 《漢語大字典》及Big5共54,711個楷體字形的構形分析及字型。
- ※ 《說文解字詁林》共11,100個小篆及重文的構形分析及字型。
- ※ 小篆和楷體的字形對映共11,056組，12,437個字形。
- ※ 《漢語大字典》的〈異體字表〉，共12,208組，包含36,309個字形。

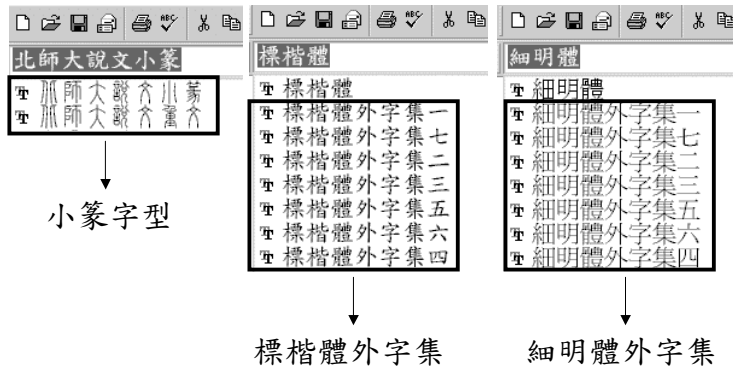
41

漢字構形資料庫 2/2

- ※ 小篆的構形分析是依據《說文解字》的解形。
- ※ 楷體字形結構的登錄是參考《康熙字典》的部首，再以橫連(△)、直連(△)及包含(△)三個法則拆分。
- ※ 楷體部件4,766個，字根1,324個。
- ※ 楷體異體字根296組，共716個字根。

42

小篆及缺字字型



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)

部件檢字

字形工具列

構形分析

異體字表

字形源流演變

45

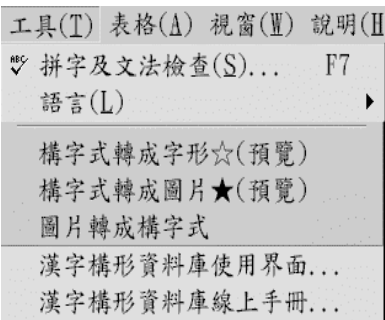
使用Microsoft Office 處理缺字

Microsoft Office缺字增益集



構字符號工具列

- 協助構字符號的輸入

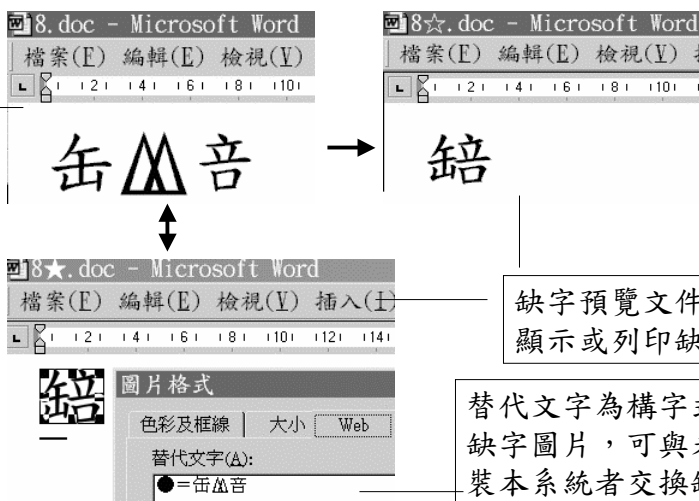


缺字次功能表

- 構字式轉成字形
- 構字式轉成圖片
- 圖片轉成構字式
- 開啟漢字構形資料庫使用界面...
- 閱讀漢字構形資料庫線上手冊...

構字式、字形及缺字圖片的轉換

編輯必須在原始文件進行



缺字預覽文件，可顯示或列印缺字。

替代文字為構字式的缺字圖片，可與未安裝本系統者交換缺字。

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	X	X	建議	部件外字
字形(型)	X	可用	優	優	不建議	字型
缺字圖片	X	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

The screenshot shows a Microsoft Word window with the following elements:

- Font Settings:** The font is set to '北師大說文小篆' (Beihua University Seal Script).
- Character Evolution (字形演變):** A window showing the evolution of the character '倍' from its components to the final form, with '說文' (Shuo Wen) as the source.
- Character Index (字形索引):** A window showing the character '倍' with its Big5 and D9AB codes highlighted.
- Font Comparison:** Three rows of the character '倍' are shown in different styles:
 - Row 1: 標楷體 (Standard Gothic)
 - Row 2: 細明體 (Thin Ming)
 - Row 3: 爪師大說文 (Seal Script)

52

使用小篆字型 2/2

編輯必須在原始文件進行

缺字預覽文件，可顯示或列印缺字。

替代文字為構字式的缺字圖片，可與未安裝本系統者交換缺字。

53

構字式反映楷體字形結構

構字式

从木，祝省聲

从水，兄聲

从心，況省聲

从示，从人、口

54

漢字構形資料庫操作界面

簡介

55

字例一：楮

字形源流演變中的小篆

字頭

楮

楮

說文·木部

楮

居延簡甲一
九九一

解形

《說文》：“楮，稅也。从木，音聲。”

(一) bǎng 《廣韻》步項切，上講並。又縛謀切。之部。

釋義

① 棍，杖。後作“棒”。《說文·木部》：“楮，稅也。”段玉裁注：“楮、棒，正俗字。”《廣雅·釋器》：“楮，杖也。”《淮南

《漢語大字典》的「楮」字

56

字形與構字式

查詢「梧」時可輸入：梧、木𣎵音、木音、木立口

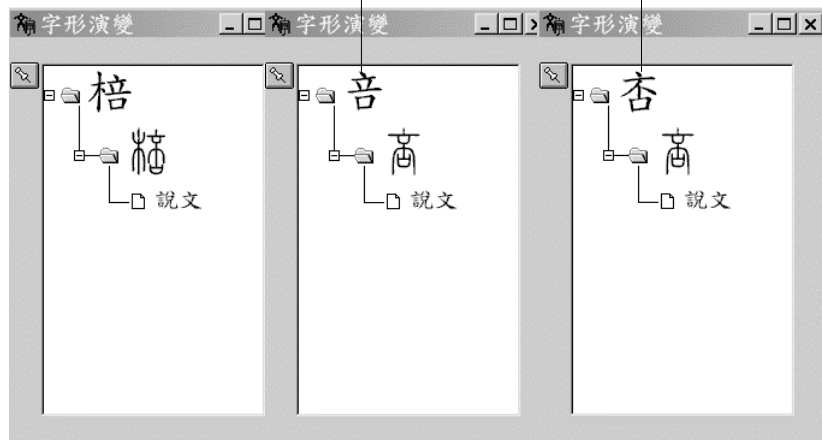


59

字形演變

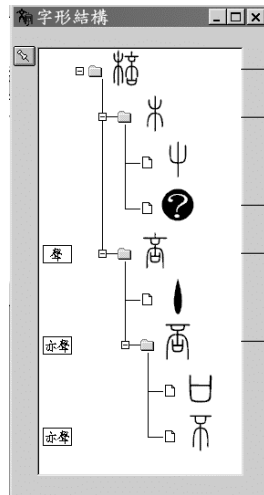
高的隸定字

高的楷化字



60

字形結構(小篆)



《說文》：梧，从木，音聲。

《說文》：木，从中，下象其根。

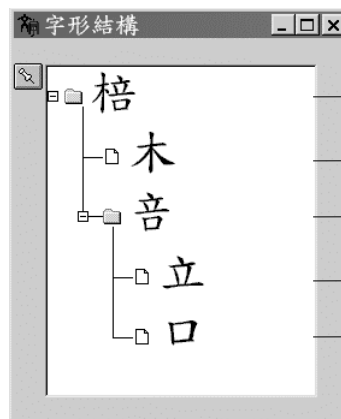
非字部件

《說文》：音，从、，从否，否亦聲。

《說文》：否，从口，从不，不亦聲。

61

字形結構(楷體)



橫向拆分(木部)

部首為字根，不再拆分。

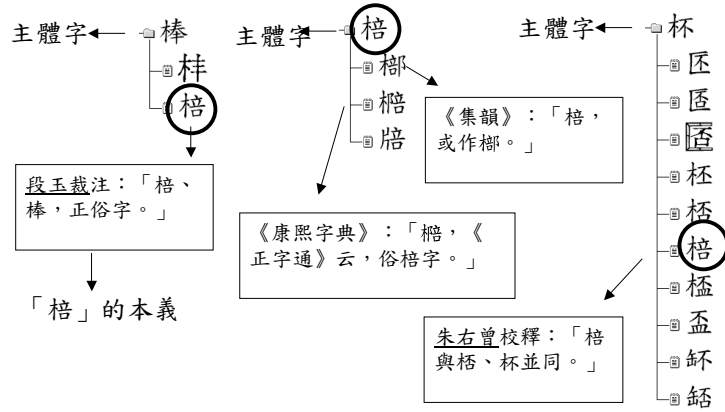
直向拆分(口部)

部首為字根，不再拆分。

部首為字根，不再拆分。

62

異體字表



<異體字表>中和「梧」相關的三組異體字

部件檢字(楷體) 1/2

依部首(康熙字典)、筆畫排序

木 3,146/54,711

- 亲
- 余
- 朶
- 榭
- 梧
- 籍
- 柚

立 489/54,711

- 乳
- 音
- 倍
- 培
- 梧
- 籍
- 敬

口 7,881/54,711

- 音
- 音
- 倍
- 培
- 梧
- 籍
- 敬

部件檢字(楷體) 2/2

依部首(康熙字典)、筆畫排序

80/54,711 37/54,711 7/54,711

65

部件檢字(小篆) 1/3

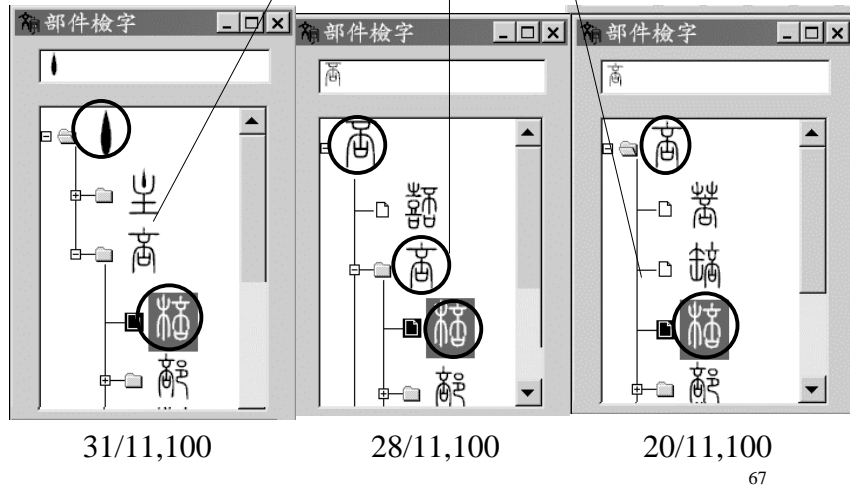
依說文解字部首排序

1,068/11,100 42/11,100 1,976/11,100

66

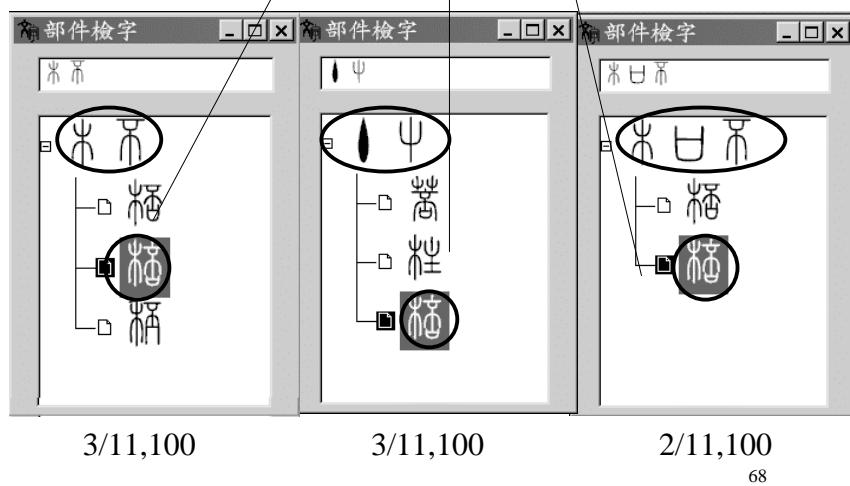
部件檢字(小篆) 2/3

依說文解字部首排序



部件檢字(小篆) 3/3

依說文解字部首排序



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

71

漢字構形資料庫操作界面

進階

72

字集功能表

楷 12 木 8 夕又√ D9AB 木卜廿口 木丛音

字集 字形 構字 選

常用字

五大碼(Big5)

漢語大字典

說文解字

以上所有楷體字

結束

部件檢字

字形結構

73

字形索引

字形索引

否

- 遠東·漢語大字典
 - 部首: 口
 - 冊-頁-字: 1-586-5
- 中文大辭典(中華學術院)
 - 部首: 口
 - 冊-編號: 2-3400
- 說文解字
 - 部首: 不
 - 說文解字詁林(鼎文書局)
 - 冊-頁: 9-950
 - 說文解字(中華書局)
 - 頁(上/下): 246下
- Unicode 3.2
 - 5426
- Big5
 - A75F

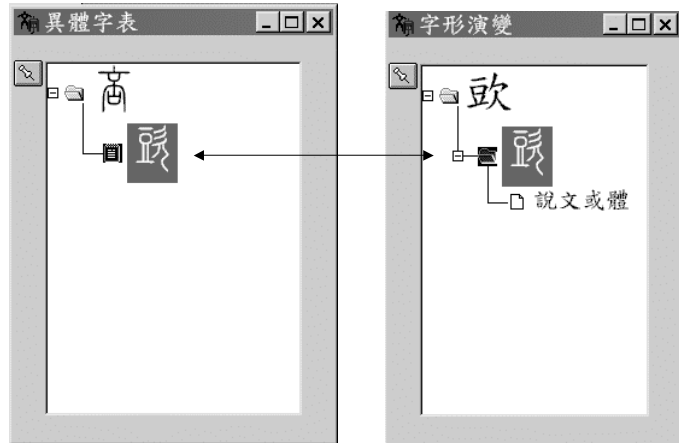
字形索引

衍

- 遠東·漢語大字典
 - 部首: 彳
 - 冊-頁-字: 2-820-4
- 中文大辭典(中華學術院)
 - 部首: 行
 - 冊-編號: 8-34854
- 說文解字
 - 部首: 水
 - 說文解字詁林(鼎文書局)
 - 冊-頁: 9-284
 - 說文解字(中華書局)
 - 頁(上/下): 229上
- Unicode 3.2
 - 884D
- Big5
 - AD6C

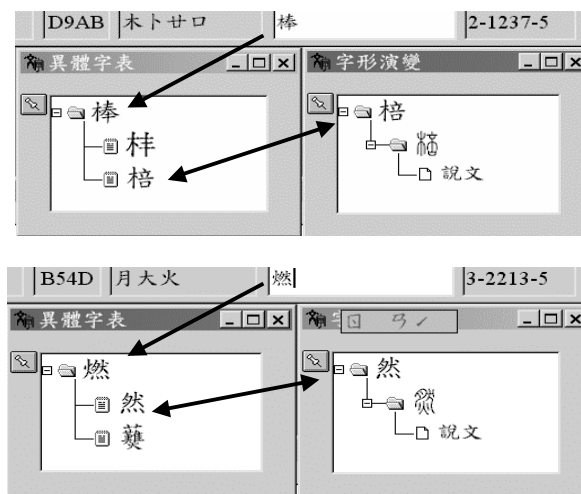
74

小篆重文



75

漢字時空座標



76

構字部件

筆畫 起筆

5 點 標楷體

康熙字典部首

- 5畫(點)
 - 玄
 - 疒
 - 穴
 - 立**

部件檢字

立

- 乳
- 音
- 倍
- 培
- 梧**
- 籊
- 啟

異體字根

異體字表

- 初
 - 壓
 - 剌**
 - 麗
 - 麗
 - 麗
 - 麗

字形演變

初

- 初
- 說文

字形演變

剌

- 剌
- 說文

異體字根

衣

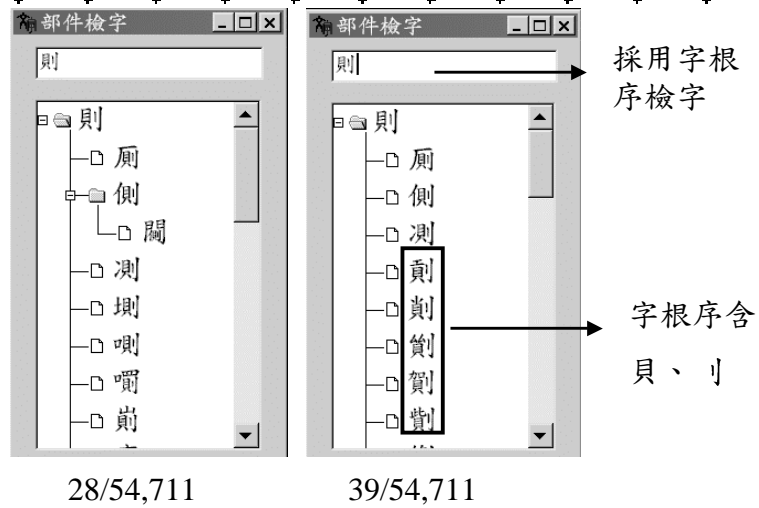
- 衣
- 衤

異體字根

刀

- 刀
- 刂
- 丩

部件檢字只列字形不列部件



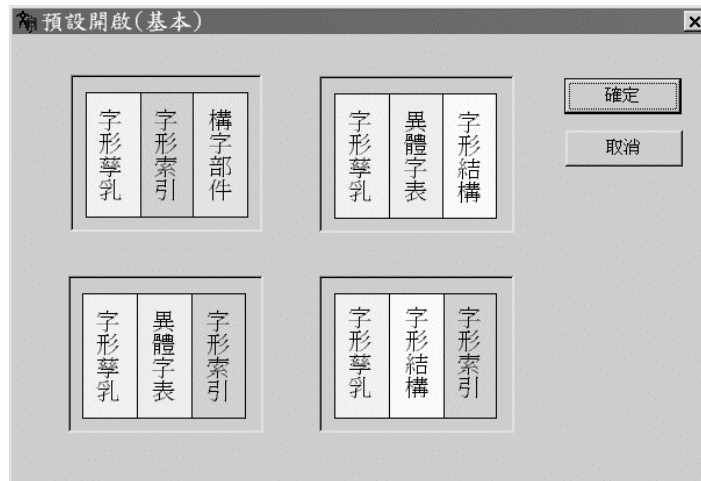
79

部件檢字包含異體字根



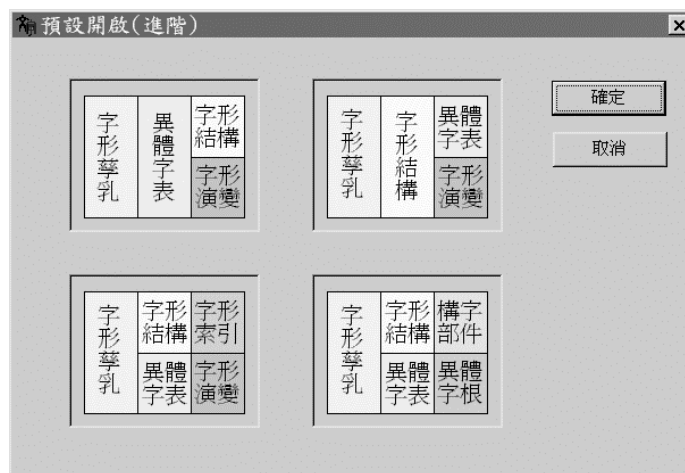
80

開啟預設視窗(基本)



83

開啟預設視窗(進階)



84

網際網路的缺字處理

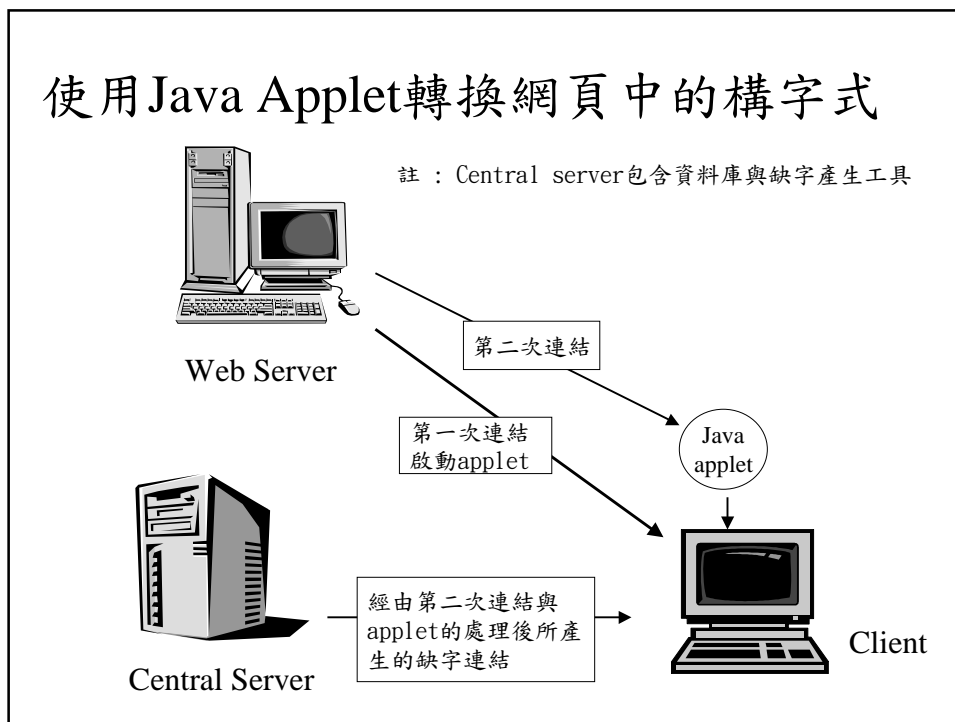
85

網頁的缺字表達

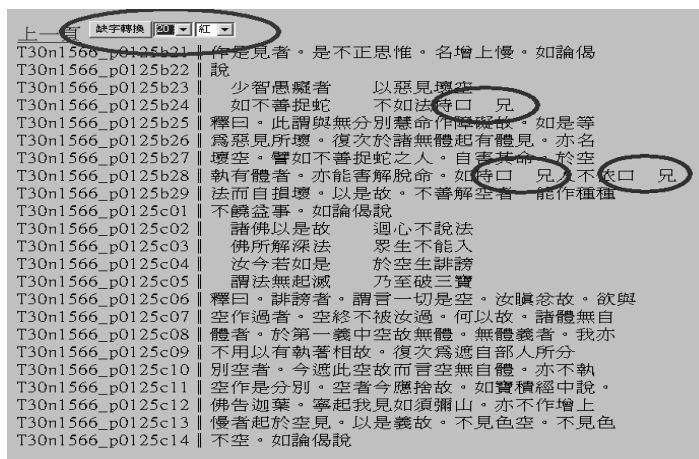
- ✱ 採用缺字圖片表達網頁的缺字。
- ✱ 瀏覽者的電腦不需額外安裝軟體或字型，即可顯示缺字。
- ✱ 瀏覽者的電腦若安裝本系統，即可將圖片轉成構字式，以作進一步的運用。
- ✱ 編製網頁時，不受字集的限制，可自由的表達缺字並與瀏覽者交換。

86

使用Java Applet轉換網頁中的構字式

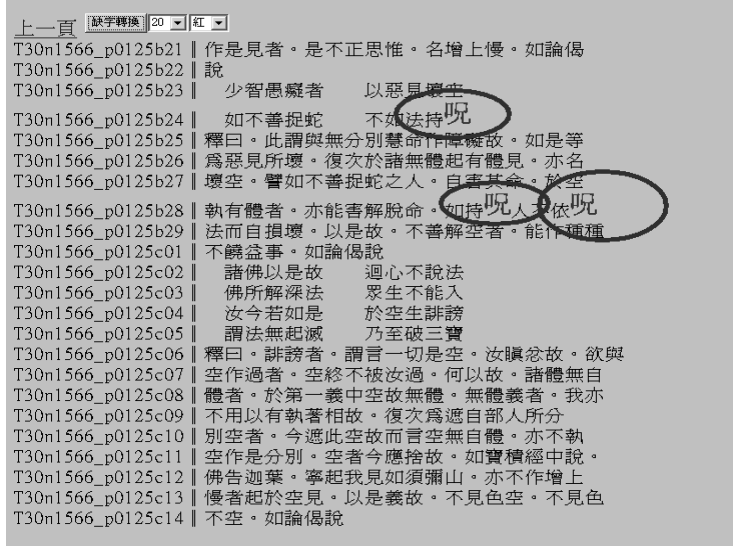


使用說明

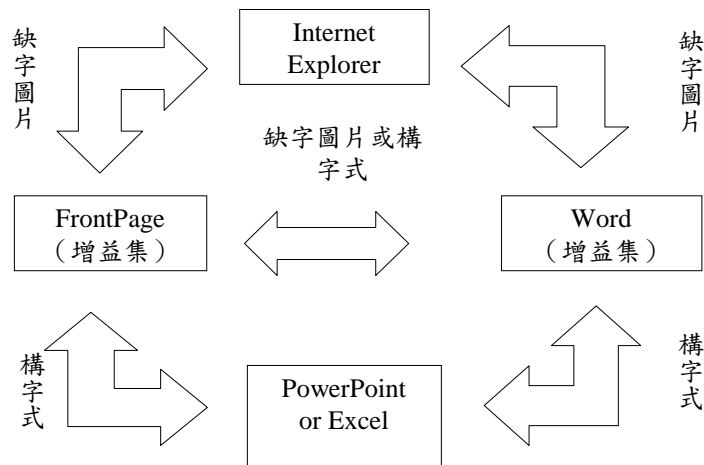


藍色所圈選之處為缺字所在，紅色圈選之處為轉換缺字的按鈕，並且可選擇所產生缺字的大小與顏色。

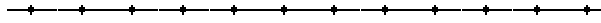
轉換後的結果,若要還原請重新載入即可



使用 Microsoft Office 處理網頁缺字

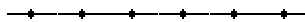
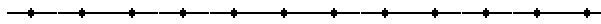


構字式的自動校正



91

使用手冊及實例操作



92

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

93

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

94

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)

95

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。

96

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。

97

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。

98

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)

99

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)

100

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)

101

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

102