

## 電子報總覽

關於本報

源碼新聞

接案 / 工作

## 專欄總覽

技術專欄

源碼秘技

法律專欄

自由專欄

企業應用

源碼專案



## 「中央研究院漢字部件檢字系統」釋出程式原始碼

建立日期 2008-04-25 08:00 最近更新在 2011-01-06 23:58

作者是 李貞臻

您在使用電腦處理中文資料時，是否常常遇到一些電腦無法輸入的文字？相關單位將漢字古籍、史料、佛典等典藏資料電腦化時，更是面臨層出不窮的缺字窘境。針對這些為數眾多的缺字，由中央研究院資訊科學研究所謝清俊研究員與莊德明助技師領導的「文獻處理實驗室」，下了多年的苦功及研究，開發了一個成本低廉又功能完備的缺字解決方案－「漢字部件檢字系統」。近日因應國內自由軟體開發的需求，宣布將其內含之「漢字部件檢字程式」，以及相關「漢字字型」以公眾授權模式釋出程式原始碼，加速「中央研究院漢字部件檢字系統」後續的推廣擴散與程式演進。「中央研究院漢字部件檢字系統」的研發最早可回溯到 1972 年交通大學關於「交大字根系統」的一系列論文和報告。1993 年，中央研究院資訊科學研究所「文獻處理實驗室」沿用「交大字根系統」的方法，進行重作改造，進而建立了「中央研究院漢字部件檢字系統」。隨著微軟作業系統不斷的改版，「漢字部件檢字系統」在這些年來陸續推出 Windows 3.1/95/98/Me/2000/XP 的版本，Vista 的版本仍在研發中。莊德明老師表示，早期由於網路不發達，所以每當參加會議發表論文，我們常分送試用磁片或光碟；而後網際網路逐漸普及，自 2002 年 10 月起，使用者已經可自行上網下載。下載人次，從早期約一天兩次，到目前一天約為十次。

整個「中央研究院漢字部件檢字系統」內含有「漢字部件檢字程式」，以及運作所需的相關「漢字字型」。在「漢字部件檢字程式」部份，考量到與其他軟體程式碼能相互融通，故採「GNU 公眾授權條款 3.0 版本〈GNU General Public License 3.0〉」進行公眾釋出；而在「漢字字型」部份，則因為其具有「圖形著作」的特性，而另行採用「GNU 自由文件授權條款 1.2 版本〈GNU Free Documentation License 1.2〉」，以及「創用 CC 姓名標示-相同方式分享台灣授權條款 2.5 版〈Creative Commons Attribution-Share Alike 2.5 Taiwan〉」兩種授權條款併行釋出。莊老師認為選擇釋出此系統的程式原始碼，不僅可以讓「中央研究院漢字部件檢字系統」嘉惠更多有需要的研究人士，更可以因為更多人的使用與回報，促進程式的改良與演化。

目前「中央研究院漢字部件檢字系統」主要的應用在於解決缺字問題，最新的版本為 2007 年 12 月上線的 2.51 版，此版收錄古今漢字 115,197 個，其中楷書字形 62,671 個，小篆及重文 11,100 個，金文 20,091 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。莊德明老師進一步指出，我們長期的目標是要解決文字學數位化的問題，所以我們正緊鑼密鼓進行 3.0 版的研發，此版最大的特色是中文字碼改用 Unicode，並可同時適用於 Windows 的繁簡體字版本。

- ◎ 「中央研究院漢字部件檢字系統」原名為「漢字構形資料庫」
- ◎ [文獻處理實驗室](#)
- ◎ [下載中央研究院漢字部件檢字系統](#)

自由軟體鑄造場電子報：第 102 期 自由軟體法律研討會菁華摘錄

分類：源碼新聞

0

[↑ Top](#)

[行動版](#)

自由軟體鑄造場 製作 最佳瀏覽狀態：IE7或Firefox2.0以上 (建議使用Firefox) · 解析度1024\*768

**E-Mail** : [contact@openfoundry.org](mailto:contact@openfoundry.org) **Address** : 台北市南港區研究院路2段128號 中央研究院資訊科學研究所 · [隱私權條款](#) · [使用條款](#) 