

# 全字庫中文網頁建構工具之研究

研究人員：潘城武、吳欣宜、馬瑪利、陳玉芳

服務機關：行政院主計處電子處理中心

## 一、研究緣起與目的

由於網際網路的蓬勃發展，使用電腦與網路的人口急速增加，依據行政院主計處電子處理資料中心九十年九出版的「台閩地區電腦應用概況報告」指出，我國個人電腦數為六百三十九萬台、上網人口已超過七百萬人，且機關企業的連網率也達到六成以上。所以不論政府、企業或個人，在建置網站並透過網路傳遞資訊的情況相當普及。但一般個人電腦作業系統多是採用 BIG-5 碼或 UNICODE 碼做為中文內碼。BIG-5 的標準常用字只有 13,053 個，而 UNICODE 也只能提供 26,183 個漢字，因此當電腦作業涉及姓名和公司行號名稱時，可能都會有無字可用的情況產生，則需連結全字庫(<http://www.cns11643.gov.tw>)或透過自行造字解決。

網頁設計者建置網頁時若用到個別電腦的自造字，使用者透過瀏覽器瀏覽網頁內容時，將無法看到完整的網頁內容，此種不便，影響中文在網際網路上的發展甚鉅，除此之外，我國目前正進行「e-Taiwan」政府旗艦計劃，為達到整體資源共享及資訊互通，不同電腦系統內碼互譯之需求日益迫切，亟需建置資訊系統的中文共同平台。本研究致力於全字庫中文網頁建構工具之探討，其以「全字庫」為基礎，開發「網頁自造字顯示工具」，解決網頁設計

者於建置網頁時中文自造字無法顯示的問題。

## 二、研究方法與過程

### (一)文獻探討

蒐集、整理與分析報章雜誌及國內外書刊文獻中有關中文網頁建構工具開發技術，並分析國內電腦中文應用環境現況，以了解本研究之研究架構、系統建置的可行性，理論探討與系統實作並行，希望在此之後，能得到一個確實可行的解決方案。

### (二)技術研討

參加國內學術界與企業界舉辦的技術研討會、造訪國內外相關技術網站，汲取新知，以瞭解目前最新資訊技術發展。

### (三)實地考察

實際參訪國內各機構，以瞭解電腦使用中文字碼所遭遇之問題並積極參與漢字編碼國際會議，使國內中文字碼標準可與國際接軌。

### (四)系統實作

本系統主要以 RedHat Linux 6.2 為作業系統，搭配 MySQL 資料庫系統，使用 Java 撰寫程式，系統架構為 Web-Base 開放式系統。採用「全字庫」做基礎，開發「網頁自造字轉換工具」，提供網頁設計者於建置網頁時

自造字得以正常顯示。

### 三、研究發現與建議

本研究藉由實作可行之工具軟體，來探討中文網頁建構工具之可行性。綜合以上各章研究結果，歸納結論與建議如下：

- (一) 本研究以建置中文電腦應用環境為目的，開發「網頁自造字轉換工具」，提供網頁建置者解決網頁自造字問題，使網頁中文字型的顯示不再受限於電腦系統本身所提供的中文字。
- (二) 本研究應用全字庫之模式，未來可用於提昇中文字碼在網際網路上的應用層次，如收集完整的中文字讀音、繁簡字形對照、詞義、多樣化字型，建置完善的中文多媒體基礎環境。
- (三) 本研究之成果可用於建置政府大型資訊系統間中文共同平台，有效的達到各系統間透過網路傳遞資訊，以達資源共享之目的，並提供政府與民間使用網路進行交換資料互通。
- (四) 本研究僅提供批次式轉換工具，未來可持續發展電腦中文字碼基礎平台，建置更具彈性的網際網路中文自造字字型顯示機制，提供使用者更多選擇。讓中文字的使用不再受限於個別作業系統的架構及字數限制。應持續透過單一窗口蒐集中文自造字，擴編國標碼，以維護國內中文資料順利交換。
- (六) 積極參與國際碼擴編，使全球漢字交換無阻礙。