

解決缺字問題 計畫書

謝清俊 2000.0509

壹、前言

對於圖書館來說，電腦缺字一直是個困擾的問題，不僅影響書目資訊的登錄、使用，更影響資訊的共享和流通。缺字問題在其他領域也是個困擾，舉凡政府的電子化、各級與遠距教學、數位化圖書館...都常為此問題傷神。

中央研究院資訊科學研究所自 1993 年起，試圖利用文字結構的知識來解決此問題。經過大量的人力投入，仔細分解各大字典和字集中十餘萬古今、正俗、繁簡等各體字形後，建立了『漢字構字資料庫』，並以此資料庫設計了解決本院『漢籍全文資料庫』缺字問題的方案。經兩年餘的測試，成效顯著。目前，除本院的資料登錄外，中華佛學研究所的大正藏佛經數位化計畫亦沿用此法。

自 1995 網際網路在國內流行以來，資訊共享之需求日益殷切；在網際網路的情境下，解決缺字問題實已到刻不容緩的地步。即僅僅為了建立網際網路上的華文內容一項，便應立刻解決缺字問題，更勿論其餘。有鑑於此，本所謹以過去所學到的經驗，提出本計畫。

貳、研究目的

- 一、研擬網際網路情境下解決缺字問題的方案，並提出雛型軟體系統，供圖書館界、教育界、政府機構...等公開試用。
- 二、本計畫以中央研究院資訊科學研究所文獻處理實驗室過去七年來處理缺字的經驗和累積的成果為基礎，研發合於國際標準的缺字表達系統和處理方式，以期能在網際網路和視窗的環境下運作。

參、時程

自中華民國八十九年六月一日起至八十九年十二月十五日止，共六個半月。

肆、研究方法與工作項目

- 一、以中央研究院資訊科學研究所以往研發的解決缺字技術為基礎。
- 二、以 BIG-5 碼為對象，整合『漢字構字資料庫』與現有的素材，發展在 WINDOW 環境下，解決缺字問題的方法。
- 三、研發合於 XML, XSL, TEI...等國際標準的缺字表達系統，以期所發展的系統適用於網際網路和各種文件處理環境。
- 四、提供使用者介面和程式介面。

- 五、整理及補充既有的字形相關資料，使之趨於完善。
- 六、如有未竟事宜，請參閱所附文件：1. 本次簡報資料,和 2.《中文字形資料庫的設計與應用》一文。

伍、預期成果

1. 本計畫的主要工作是系統整合：將既有的各個子系統組合成為一個可裝設在 Window 下和在網際網路上使用的系統。
2. 此系統之設計思考與所有的交換碼完全不同，是故可與任何交換碼相容。唯一的需求是原交換碼必須提供約一千個使用者編碼空間。
3. 本計畫所產出的系統有以下的優點：
 - 解決或大量舒緩缺字所引起的下列問題。
 - 資料登錄工作時造字的管理不易。
 - 交換碼使用者造字碼位的空間不足。
 - 異體字造成文件檢索和處理上的困擾。
 - 由於自行造字造成的資訊共享障礙。
 - 可擴大現行碼的空間而無需增估碼位。
 - 可應付絕大部份未來遭遇到新字的問題而不需更動系統結構。
 - 可供作為中文資訊交換之基礎。
 - 本計畫產出所用的標誌語言與國際標準相容，故可以適用於各種數位化的文獻工程而不致於產生文化上的障礙。
 - 採用本計畫產出的系統，可以避免字集加大時引起的字碼加長，而必需修改或重寫應用程式的麻煩。
4. 本計畫願將產出之系統免費提供國人使用。本所文獻處理實驗室現有的產出可以免費提供者包括：
 - 漢字構形資料庫
 - 普及組字根集(801個，含字型和倉頡輸入碼)
 - 部件集(2600個含字型和倉頡輸入碼)
 - 構字式集(56906個字的構形)
 - 使用者界面和程式界面
5. 提供系統光碟、平實易讀的使用手冊和安裝手冊。
6. 本系統可作為發展文字學資料庫的基礎。

陸、預算

預算項目	申請金額	說明
一、人事費	335,500	<ul style="list-style-type: none"> • 研究主持人1名，月支10,000元(6.5月計) • 協同主持人1名，月支8,000元(6.5月計) • 兼任研究員2名，8,000x2=16,000(6.5月計) • 按時計酬臨時人員:800元/日 x(校對資料80日) 小計64,000 • 出席費
二、業務費	300,000	<ul style="list-style-type: none"> • 印刷費、文具紙張費 (出版光碟、使用手冊、系統安裝說明) • 稿費及審稿 • 軟體設計費 • 資料蒐集費 • 電子計算機使用費 • 耗材費
三、旅運費	88,000	<ul style="list-style-type: none"> • 聘請國外文獻工程及文獻表達標準之專家一人來工作一星期，機票60,000、生活費4,000x7天 計88,000
四、雜支	30,000	國內外郵電費
五、行政管理費	30,000	
合計	783,500	

註：為了要使缺字的表達方式能夠上網際網路，能夠與文字處理及文件處理(包括文件檢索)的軟體相容，我們需要聘請懂得 XML、XSL、TEI 等標準且有設計經驗的國外專家一名，在此工作一週。目前此位專家為資深教授 Professor John Lehman 任職於阿拉斯加大學，此經費係依該地來台之差旅費預估。此項聘請國外專家之工作極為重要，否則本計畫將失去申請之原意。