



Unicode工作小組 工作成果及現況

報告人：顧力仁主任(國家圖書館書目資訊中心)

Innopac/Millennium台灣地區使用者聯盟會議

@國立中山大學圖書館

2005.12.12



Unicode工作小組圖書館成員

- 宋玉先生（國圖自動化技術顧問、中文字集字碼專家）
- 中央研究院（計算中心、近史所圖書館）
- 國立中山大學圖書館（採編）
- 國立中央圖書館台灣分館（閱覽）
- 國立成功大學圖書館（採編）
- 國立政治大學圖書館（採錄、編目、系統）
- 國立臺灣大學圖書館（編目、系統資訊）
- 國家圖書館（書目中心、編目、研究、資訊）
- 逢甲大學圖書館（技術服務）
- 朝陽科技大學圖書館（採訪編目）
- 臺北醫學大學圖書館（編目）
- 輔仁大學圖書館（編目、資訊）



Unicode工作小組廠商成員

- Innovative Interfaces Inc.
- 澳大利亞商美科資訊股份有限公司
台灣分公司
- 傳技資訊股份有限公司
- 鼎盛資科股份有限公司



主要工作計畫

- 蒐集、編製、維護字碼對照表
 - CCCII_Unicode對照表
 - 多碼對一字對應表
 - 同義異體字參照對應表
- 公布最新版對照表
- 缺字蒐集、彙整、通報
- Unicode工作小組網站建置與維護



分工合作方式

- 字碼表來源：宋玉先生、廠商、專家
- 字碼表彙整/編製對應表初稿：國家圖書館
- 對應表確認：工作小組成員
- 對應表修訂：工作小組成員
- 相關程式開發：國家圖書館
- 網站建置維護：國家圖書館、輔仁大學圖書館



第一階段工作 & 成果

1. 蒐集現有CCCII_Unicode對照表

※來源：宋玉先生、Innovative Interfaces Inc.、臺灣大學圖書館。

2. 彙整不同來源對照表

- (1) 訂定彙整原則；
- (2) 撰寫彙整程式；
- (3) 匯集於Access資料庫；
- (4) 編製雙向對照表。



第一階段工作&成果（續）

4. 問題討論及解決方式
 - ※問題—CCCII同形異碼字造成資料交換及字碼轉換問題。
 - ※解決—優先處理同形異碼的對應問題。
5. 從CCCII-Unicode對照表中擷取多對一字集，討論決定Unicode轉CCCII之選碼原則。
6. 小組成員分工進行人工查核多對一對照表。



第一階段工作 & 成果 (續)

7. 成果—

- 1) CCCII_Unicode對照表計27,436碼。
- 2) Unicode_CCCII對照表計25,868碼。
- 3) 多碼對一字形對照表計2,906個CCCII碼
對應1,325個Unicode字形。

8. 對照表推廣利用

9. Unicode 4.0新增字碼對照前置作業。



第二階段工作 & 成果

1. 宋顧問提供Unicode新增字碼與CCCII對照表初稿，由工作小組分工查核。
2. 傳技、鼎盛公司提供CCCII_Unicode對照表，與第一階段對照表進行比對驗證，增補部份闕漏字碼，並提供廠商修訂。
3. 蒐集NBINet聯合目錄資料庫所用之CCCII碼，與彙整後之對照表比對，過濾出尚未對應Unicode之CCCII。



第二階段工作 & 成果 (續)

4. 國家圖書館先行查找新增對應字碼，再由工作小組分工查核。
5. 成果—
 - 1) CI_UNI_V2(940801)對照表計41,513碼。
 - 2) UNI_CI_V2(940801)對照表計39,364碼。
6. 輔仁大學圖書館及國家圖書館共同建置Unicode工作小組網站(<http://unicode.ncl.edu.tw>)，提供對照表下載並分享相關資訊。

Unicode工作小組網站

Unicode工作小組 - Microsoft Internet Explorer

檔案(E) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

← 上一頁 → 搜尋 我的最愛 媒體

網址(D) <http://unicode.ncl.edu.tw> 移至 連結 >>

Search Web Mail My Yahoo! Games Personals Music >>

國家圖書館
National Central Library

關於Unicode小組

小組成員一覽

字碼對照查詢

歷次會議重要決議

成果報告

Download

回首頁

Unicode工作小組

Unicode Work Group

【最新消息】

最新版CCCII與Unicode對照表「CI_UNI_V2(940801)」已公布，對照表包括兩部分：(1)可對應Unicode之41,513個CCCII碼；(2)可對應CCCII之39,364個Unicode碼。歡迎各界下載利用，並提供修訂建議。

國家圖書館著作權聲明
Copyright©2004.
National Central Library.
All rights reserved
洽詢專線：(02)23822374

完成 網際網路

<http://unicode.ncl.edu.tw>



參考資料

- 中文資訊交換碼/行政院文化建設委員會資訊應用國字整理小組編；張仲陶主編. 第二冊第三版. 臺北市:行政院文化建設委員會,民74[1985](民81,重印)
- 中文資訊交換碼/行政院文化建設委員會資訊應用國字整理小組編；張仲陶主編. 第三冊. 臺北市:行政院文化建設委員會,民76.
- 中文資訊交換碼異體字表/資訊應用國字編委會編. 第一次綜合修訂稿. 臺北市:行政院文建會,民81.
- The Unicode standard, version 4.0 : the Unicode consortium /edited by Joan Aliprand ... [et al.]. Boston, MA : Addison-Wesley, c2003.
- Unihan Database (<http://www.unicode.org/charts/unihansearch.html>)
- CNS11643中文標準交換碼全字庫 (<http://www.cns11643.gov.tw>)
- 運用PN300查核昌泰中文內碼



問題探討

- 圖書館在各時期使用不同之中文系統（MS Windows、永麒、昌泰...），造成同一字形前後使用字碼不一致情形。
- 字碼表之外的使用情形難以掌握，無法完整蒐集所有字碼。
- Unicode仍在修訂中，CCCII_Unicode對照表須配合修訂。



未來工作計畫

- 進行對照表後續之增修
- 同義異形字參照對應表
- 缺字蒐集及後續處理



CCCIUnicode對照表後續修訂

- 蒐集圖書館界提供之闕漏字形字碼。
- 配合Unicode新版之增訂狀況進行增修。
(本年度新增對應之CCCI預計將有10,000餘字碼，其中大多為異體字。)
- 蒐集使用者提出之建議進行修訂。
- 不定期公佈勘誤表。
- 視修訂幅度公佈新版對照表。



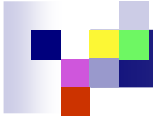
同義異形字參照對應表

- 工作小組成員持續蒐集同義異形之字碼及字形。
- 訂定同義異形字之認同原則，並進行認同原則之推廣。
- 建立同義異形字參照對應表，作為各系統提供同義異形字對照查詢之依據。
(註：並非所有系統皆有此功能)



缺字蒐集及後續處理

- 蒐集各館提供至國家圖書館之缺字相關資料須包括：缺字字形、書頁影本、初步處理說明、館別、該筆紀錄在該館資料庫之紀錄號碼。
- 訂定缺字蒐集、提報程序及負責單位。
- 訂定圖書館界缺字暫時處理方式。



敬請指教
謝謝！

