

去年底我與 **suns99** 兄初步完成了《教育部異體字字典》宋體字頭的清理工作後（尚有 **13830** 個手寫字形待清理），本來打算休息一年，先把工作重點放在拆分數據的完善上。結果阿文兄硬是不讓我休息，跳出來說要幫忙清理手寫字形的部份，呵呵！計畫趕不上變化，於是只好牙根一咬繼續拼下去了。在經過波波折折、兩度重新編碼後，終於在半年後完成了全部的清理工作，共新增 **8289** 個手寫字形至字庫，《教育部異體字字典》的所有字頭全數「文字化」。至此「全宋體」這個大型字庫，收字正式突破了十八萬漢字，涵蓋了四大字典（《漢字海》、《教育部異體字字典》、《中華字海》、《漢語大字典》）的所有字頭，應該足供專業等級的漢字應用。

核心版本：0.9.3.2 (2022年6月)
發布網址：[附件檔案](#) (WFG製作)

收字的工作雖然「大勢底定」，但後續尚有兩、三萬字的拆分數據需要優化，只能繼續「長期抗戰」。先將這「並不完善」的「全宋體」字庫初稿發布出來，讓大家能先使用，也為這波大型字典的清理工作正式做一個 Ending。

《教育部異體字字典》的字頭全數「文字化」，對我來說是一個重要的里程碑。盼了多年，沒想到真有實現的一天，到現在還有些微「美夢成真」的不真實感。這些《異體字字典》的字頭，對於古籍的數位化具有相當程度的幫助，可以大幅減少缺字的機會，有利於加速古籍的數位化整理，希望有更多的「有志之士」能投身到這個領域，將這些資源進一步地發揚光大。

由於全數「文字化」了，《教育部異體字索引字典》已經不再需要字頭圖片，所以原先的 mdd 檔已經不再需要，換上新版之後便可刪去（[《教育部異體字索引字典》](#) 另文更新）。

很多字形相同的手寫字在官網上是以各自獨立的圖片呈現，彼此並不關聯。這些重複的手寫字，在整理的過程中都已經對應成同一字，也就是說它們都被關聯起來了，使用者查詢時，可以得到更全面的資訊，不會如官網般只能查得局部。《教育部異體字字典》現有字頭共 106315 字，若將同形字合併，實際收字 96800 字，略少於《漢字海》。

感謝阿文兄、suns99 兄，沒有他們的話，這失落的「最後一塊拼圖」也不可能這麼快就拼上去。在這裡還要感謝另一位好友 Masta meta，陸續提供了一些他手工精造的字形讓我替換掉品質較差的字，雖然數量還不多，但至少是慢慢走在優化的道路上，呵呵！

樂見學術研究、教育工作、個人閱讀這方面的運用，但請勿用做任何形式的商業營利行為。希望「全宋體」這個大型字庫以及「部件檢索」這個檢字工具，能在漢字文化的整理、研究上幫上一點小忙。

下載連結：[倉頡碼表.7z](#) (僅保留漢字部分，請自行併入您慣用的碼表)

最後，將一些整理過程中記錄的工作日誌附在這裡，做為一個回憶與紀念：

- 2021/12/31 阿文兄留言表示願意幫忙清理《異體字字典》手寫字頭。
- 2022/01/03 將欲清理的 13826 個手寫字製成的臨時性字型檔及字頭表列檔發給阿文兄，正式啟動作業。
- 2022/02/10 阿文兄完成清理，共還原 4661 字，還原率約為 34%。
- 2022/04/06 結束位於 15 字面的《異體字字典》宋體字頭之拆分數據優化工作，開始處理手寫字頭。
- 2022/05/06 阿文兄的還原字複驗完成，累計剔除不正確 482 字，故實際的有效還原為 4179 字，還原率約為 30%。
- 2022/05/07 另行還原阿文兄未檢出的標準字 313 字，剔除重複字 55 字。
- 2022/05/13 手寫字頭編碼完成。
- 2022/05/16 更新全宋體、部件檢索及《異體字索引字典》，共 9279 個手寫字頭加進補充字，預備發布。發現阿文兄的失誤率有點高，suns99 兄開始進行複查，暫時取消發布。
- 2022/06/06 suns99 兄複查完畢，9279 個手寫字中又還原了 990 字。
- 2021/06/07 針對 suns99 兄的還原複驗完成。在兩位好友的協助下，實際的有效還原為 5169 字，還原率約為 37%。
- 2021/06/08 手寫字頭重新編碼完成。更新全宋體、部件檢索及《異體字索引字典》，共 8289 個手寫字頭加進補充字，預備發布。

勘誤：