

中華民國國家標準	中文標準交換碼使用方法	總號	11643-1
CNS		類號	X5012-1

The Usage of Chinese Standard Interchange Code

1. 適用範圍：本標準適用於中文標準交換碼之使用方法，並介紹 CNS 5205〔資訊處理及交換用七位元碼字元集（組）〕、CNS 7654〔資訊處理—七位元及八位元集—延碼技術〕等相關標準以爲參考。
2. 用語釋義：依 CNS 5205、CNS 7654 及 CNS 11643〔中文標準交換碼〕之資訊用語釋義解釋如下，以協助使用者了解本使用說明。
 - (1) 位元組合 (bit combination)：爲一有序之位元編組亦即爲一字碼，用以代表一字元。在七位元環境中，位元組合後之碼的範圍爲 00 ~ 7F，在八位元環境中則爲 00 ~ FF。
 - (2) 字元 (character)：爲一個基本位元組合，用於資料之組織、控制或表示，亦有人翻譯爲字符。在 CNS 11643 中即代表一表意文字 (Ideograph Character)，亦即爲中文字。
 - (3) 碼字元集 (coded character set)：爲一有明確法則建立之字元集，其中每一字元及其位元組合間有一對一之對應關係。
 - (4) 延碼 (code extension)：將未包含於現有碼字元集中之字元加以擴充編碼之技術，即字集擴充之編碼技術。
 - (5) 碼表 (code table)：說明碼中每一字元對應於每一位元組合（字碼）之表。
 - (6) 控制字元 (control character)：係由單一位元組合所組成之碼化表示法，表相對應之控制功能。
 - (7) 控制功能 (control function)：係能影響資料記錄、處理、傳輸或解釋之功能，可由一個或一個以上位元組合所組成之碼化表示法來表示。
 - (8) 圖形字元 (graphic character)：係控制功能以外之字元，通常代表目視可辨之手寫、列印或顯示。
 - (9) 位置 (position)：在碼表上之某個項目，以行和列之座標表示。
 - (10) 指定 (to designate)：在規定之方法下，指定使用字元集，此字元集可立即出現，也可跟隨在一些控制功能之後出現。
 - (11) 設備狀態 (environment)：在資訊處理系統、資訊交換系統或其他類似之系統設備中，字元所使用之位元數目特性。如七位元設備狀態、八位元設備狀態等。
 - (12) 逸出順序 (escape sequence)：在延碼過程中達成控制作用之一串位元，由兩個或兩個以上之位元組合所組成，其第一個位元組合對應於“逸出”(ESC)字元。
 - (13) 終結字元 (final character)：使逸出順序終結之字元。
 - (14) 中間字元 (intermediate character)：在包含兩個字元以上之逸出順序中，出現於逸出字元與終結字元間之字元。
 - (15) 圖形符號 (graphic symbol)：圖形字元或控制功能之目視表示。
 - (16) 挑取 (to invoke)：當一規定之位元組合出現時，能使一已經指定之字元集爲此位元組合所使用，直到適當之延碼功能發生。
 - (17) 表示 (to represent)：用一規定之位元組合，作爲已被指定及挑取之字元集中之字元；或者以一逸出順序作爲另加控制字元之用。

(共 30 頁)

公布日期
84年1月4日

經濟部標準檢驗局印行

修訂日期
年 月 日

印行日期94年10月

本標準非經本局同意不得翻印

A4 (210 × 297)

3. 編碼原則：CNS 11643 係依據 CNS 7654 第 5.1.3 節之多位元組字元集之規範而訂定。採用兩位元組之編碼架構，每一位元組之編碼範圍為 21-7E 共 94 個編碼位置，兩個位元組即 94*94 共 8,836 個編碼位置，即為一字面或一字元集。為考慮大型資訊系統之需求及參考 CNS 7654 裡有關專用終結字元之規定（可以有 16 個字面），CNS 11643 之架構暫定 16 個字面，其中 1 ~ 11 字面為編列國家標準用，12 ~ 16 字面則為用戶自訂區，使用者可視應用需求來編列特殊之字元。

依據 CNS 7654 七位元之延碼技術，每一字面可利用逸出順序字元碼及鎖定移位功能控制碼來指定並選用字元。每一字面並得依規定向國際標準組織 (ISO) 申請終結字元 (Final Character)，以組成逸出順序字元碼指定字面。故為提高傳輸效率，每一中文字元應依其使用頻率及使用範疇加以編碼，使常會一起出現的字編在同一字面中，減少字面轉換之頻率。同時，為使得每一字面之排序有所依據，每一字面之字集以筆劃部首為排序依據。

4. 字集編排原則：CNS 11643 國家標準之字集範圍原則上以教育部所公佈之「常用國字標準字體表」、「次常用國字標準字體表」、「罕用字字表」及「異體國字字表」為主要參考依據。但因應用需求而不包含於上列四個字表內之字集，則編排於使用者加字區即由第 12 ~ 16 字面內。另為提高傳輸效率，將字集依使用頻率及範疇分別編列於第 1 ~ 7 字面，而戶政系統所須用到之特殊字元則編列於第 15 字面。各字面字集之編列現況說明如下。

第一字面之字集：本編碼系統為減少字面轉換次數，特編最常用之中文字於交換碼第一字面，所編字彙除教育部頒訂之「常用國字標準字體表」內所列 4,808 字外，並蒐集國內民國 71 年時之國中、國小教科書之常用字 587 字及異體字 6 字，共收 5,401 字，字碼區間為 4421 至 7D4B。

第二字面之字集：第二字面之字集共收 7,650 字，從教育部所頒「次常用國字標準字體表」及另外蒐集罕用字表中使用頻率較高之中文字 1,320 字而來。字碼區間為 2121 至 7244。

第三字面之字集：第三字面之字集為 77 年 6 月由行政院主計處電子處理資料中心所整理出來，包括教育部罕用字及異體字，原放置於使用者加字區第 14 字面之第一部份 6,148 字之字集及字序。字碼區間為 2121 至 6246。

第四字面之字集：第四字面之字集為從戶政用字，資訊技術—廣用多八位元編碼字元集第二版之漢字，81 年時之各單位／資訊業等用字及原第 14 字面之第二部份 171 字整理而成，共 7,298 字。字碼區間為 2121 至 6E5C。

第五字面之字集：第五字面之字集為教育部所頒布之罕用字，而不包含於前 4 個字面之字集，共 8,603 個字。字碼區間為 2121 至 7C51。

第六字面之字集：第六字面之字集為教育部所頒布之異體字而不包含於前 5 個字面，且其筆畫在 14 畫（含）以下之字集，共 6,388 個字。字碼區間為 2121 至 647A。

第七字面之字集：第七字面之字集為教育部所頒布之異體字而不包含於前 6 個字面之字集，共 6,539 個字。字碼區間為 2121 至 6655。

第一字面除包含 5,401 個常用字外，並包含了 684 個符號及文字如下。

- (1) 間隔符號 1 個。
- (2) 標點符號 28 個。
- (3) 括號及製表符號 89 個。
- (4) 一般符號 34 個。
- (5) 學術符號 51 個。
- (6) 單位符號 31 個。
- (7) 數字符號 42 個，包括阿拉伯數字 10 個、羅馬數字大小寫共 20 個、中國數字 12 個。

(8) 外文字母 100 個，包括大寫英文字母、小寫英文字母各 26 個、大寫希臘字母、小寫希臘字母各 24 個。

(9) 國語注音符號 42 個。

(10) 數字序列符號 20 個。

(11) 中國文字部首 213 個，「久久」兩部首同歸於「久」部中。

備註：「久」部首尚未編碼，得將來擴編時，一併列入修訂。

(12) 控制碼圖形符號 33 個。

5. 中文碼使用方式：中文資訊處理主要是要利用資訊技術來處理中文資料，包含了中文資料的輸出入及處理。而為了方便電腦處理中文資料，我們必須給每個中文字有其固定代碼以資區別，就像每個人有其自己的身份證字號。這些代碼主要又可分為輸入碼、內碼及交換碼。其中輸入碼又隨各種不同的輸入法及不同的考慮觀點，而有不同的代碼，這些不在本文中討論。

基本上交換碼之目標在確保資訊之正確交換，故各國政府均依據自己國家文化的需求而制訂資訊交換字元集 (Character Set for Information Interchange) 之國家標準，除規定交換碼之標準外，並規定了每個字集之順序。而內碼除須考慮所需之字集大小外，尚須考慮應用程式之可攜性及處理效率。

如果交換碼與內碼相同，則當然是最好的安排，如 ASCII(American Standard Code for Information Interchange) 除了是交換碼外，也被拿來當內碼使用。但表意文字由於字數較多，並且依據 CNS 7654 之規定，則其中會夾雜著字面轉換及移位功能之控制碼，變成非定長度，除了影響處理效率外，並降低了應用程式之可攜性。因此一般內碼架構均參考國家標準之字集及字序及原作業系統之編碼架構，進行漢字編碼之擴充，如 BIG-5 碼、TCA 碼、SOPS 碼及 EUC 碼等內碼架構（如附錄一所示）。因此在電腦內部處理時，使用內碼即可，而在傳輸時，再轉換為標準交換碼傳輸。

6. 字集使用原則：當需求字集越多時，其所需內碼架構愈大，即每個字元須更多位元組來表示，故所佔之記憶體空間也愈大，除影響處理效率外，也負擔了更大之成本，因此每個使用者應選擇適當之字集。

依 CNS 11643 之內容，第一、二字面之字集已包含了一般常用之中文字元，應可滿足一般之財務、文書等應用系統。但對於較大之資訊處理系統，則依其應用性質，參考第三、四字面之字集說明，考慮包含第三、四字面之字集。由於我國規定人名可自康熙字典中選取所需之字，故以處理人名為主之戶政系統等，除須包含第 15 字面之人名專用字集外，亦可考慮包含 5~7 字面之字集。

原則上使用者應依應用需求，選擇適當之字集，而非字集越多越好。至於少數之特殊文字及符號，則可考慮先自行編入使用者加字區使用。

7. 字集之指定與字元之擷取：依 CNS 7654 之規定，CNS 11643 的每一字面已向國際標準組織申請登記，並由國際標準組織對每個字面給予一相對應之終結字元。目前 CNS 11643 之一、二字面已向國際標準組織完成登記，並獲得相對應之終結字元 4/7 及 4/8，而三至七字面亦正在申請中。有關各字面與終結字元之關係如下表所示。