﻿

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 請於空白格上按滑鼠右鍵選取輸入法，拼成並選字後按[Enter]鍵與[查詢][>使用說明](http://www.cns11643.gov.tw/AIDB/file.do?path=files/others/java_swing.pdf) | |  |  |

﻿

* + [全字庫介紹](http://www.cns11643.gov.tw/AIDB/intro.do)
  + [全字庫現況](http://www.cns11643.gov.tw/AIDB/statistics.do)
  + [中文碼介紹](http://docs.google.com/CNS11643%20中文全字庫-認識全字庫-中文碼介紹_files/CNS11643%20中文全字庫-認識全字庫-中文碼介紹.htm)
  + [專有名詞介紹](http://www.cns11643.gov.tw/AIDB/terms.do)
  + [全字庫使用說明](http://www.cns11643.gov.tw/AIDB/usages.do)
  + [線上教學](http://www.cns11643.gov.tw/AIDB/tutorial.do)
  + [相關網站](http://www.cns11643.gov.tw/AIDB/links.do)
  + [國際字碼相關組織](http://www.cns11643.gov.tw/AIDB/organizations.do)
  + [全字庫授權聲明](http://www.cns11643.gov.tw/AIDB/copyright.do)
* + [複合查詢](http://www.cns11643.gov.tw/AIDB/query_composite.do)
  + [筆劃查詢](http://www.cns11643.gov.tw/AIDB/query_general.do?active=0)
  + [注音查詢](http://www.cns11643.gov.tw/AIDB/query_general.do?active=1)
  + [部首查詢](http://www.cns11643.gov.tw/AIDB/query_general.do?active=2)
  + [倉頡查詢](http://www.cns11643.gov.tw/AIDB/query_general.do?active=3)
  + [CNS查詢](http://www.cns11643.gov.tw/AIDB/query_general.do?active=4)
  + [Big5查詢](http://www.cns11643.gov.tw/AIDB/query_general.do?active=5)
  + [Unicode查詢](http://www.cns11643.gov.tw/AIDB/query_general.do?active=6)
  + [拼音查詢](http://www.cns11643.gov.tw/AIDB/query_general.do?active=7)
  + [符號查詢](http://www.cns11643.gov.tw/AIDB/query_general.do?active=8)
  + [拼音文字查詢](http://www.cns11643.gov.tw/AIDB/query_general.do?active=9)
  + [部件查詢](http://www.cns11643.gov.tw/AIDB/query_general.do?active=10)
  + [筆順序查詢](http://www.cns11643.gov.tw/AIDB/query_general.do?active=11)
* + [碼表查詢](http://gateway.cns11643.gov.tw/AIDB/codetable_search.do)
  + [文件轉碼](http://gateway.cns11643.gov.tw/AIDB/NewtransFile.do)
  + [轉碼 Web Service](http://www.cns11643.gov.tw/AIDB/CNSTransferInfo.do)
  + [程式下載](http://gateway.cns11643.gov.tw/AIDB/download.do?name=%E5%80%8B%E4%BA%BA%E9%9B%BB%E8%85%A6%E9%80%A0%E5%AD%97%E8%99%95%E7%90%86%E5%B7%A5%E5%85%B7%3EUnicode%E5%B9%B3%E5%8F%B0)
* + [新字申請處理流程](http://petition.cns11643.gov.tw/AIDB/words_flow.do)
* + 全字庫授權
    - [授權聲明](http://www.cns11643.gov.tw/AIDB/copyright.do)
    - [授權下載](http://www.cns11643.gov.tw/AIDB/cns_authorization_apply.do)
  + [全字庫線上編輯器](http://www.cns11643.gov.tw/AIDB/cns_editor.do)
  + [活動快訊](http://www.cns11643.gov.tw/AIDB/cns_activities.do)
  + 其他
    - [小遊戲](http://www.cns11643.gov.tw/AIDB/cns_games.do)
    - [桌布下載](http://www.cns11643.gov.tw/AIDB/cns_wallpaper.do)
* + [中文共通平台](http://www.cns11643.gov.tw/AIDB/cns_javaswing.do)
  + [字型即時顯示](http://www.cns11643.gov.tw/AIDB/cns_font.do)
  + [造字解決方案](http://www.cns11643.gov.tw/AIDB/cns_create_word.do)
* + 個人電腦造字處理工具
    - [Unicode平台](http://www.cns11643.gov.tw/AIDB/download.do?name=%E5%80%8B%E4%BA%BA%E9%9B%BB%E8%85%A6%E9%80%A0%E5%AD%97%E8%99%95%E7%90%86%E5%B7%A5%E5%85%B7%3EUnicode%E5%B9%B3%E5%8F%B0)
    - [Big5平台](http://www.cns11643.gov.tw/AIDB/download.do?name=%E5%80%8B%E4%BA%BA%E9%9B%BB%E8%85%A6%E9%80%A0%E5%AD%97%E8%99%95%E7%90%86%E5%B7%A5%E5%85%B7%3EBig5%E5%B9%B3%E5%8F%B0)
  + [字型下載](http://www.cns11643.gov.tw/AIDB/download.do?name=%E5%AD%97%E5%9E%8B%E4%B8%8B%E8%BC%89)
  + [文字編輯工具](http://www.cns11643.gov.tw/AIDB/download.do?name=%E6%96%87%E5%AD%97%E7%B7%A8%E8%BC%AF%E5%B7%A5%E5%85%B7)
  + [輸入法](http://www.cns11643.gov.tw/AIDB/download.do?name=%E8%BC%B8%E5%85%A5%E6%B3%95)
  + [PDA工具](http://www.cns11643.gov.tw/AIDB/download.do?name=PDA%E5%B7%A5%E5%85%B7)
* + [問題集](http://www.cns11643.gov.tw/AIDB/service_faq.do)
  + [客服諮詢](http://www.cns11643.gov.tw/AIDB/service_consulting.do)

﻿

首頁 › 認識全字庫 › 中文碼介紹

### 中文碼介紹

#### 相關之中文碼

* [CNS11643 國家中文標準交換碼](http://www.cns11643.gov.tw/AIDB/encodings.do#encode1)
* [BIG-5 碼](http://www.cns11643.gov.tw/AIDB/encodings.do#encode4)
* [BIG-5E 碼](http://www.cns11643.gov.tw/AIDB/encodings.do#encode5)
* [BIG-5碼香港增補字符集(zh\_HK.hkbig5)](http://www.cns11643.gov.tw/AIDB/encodings.do#encode6)
* [EUC 碼](http://www.cns11643.gov.tw/AIDB/encodings.do#encode7)
* [CCCII 碼](http://www.cns11643.gov.tw/AIDB/encodings.do#encode8)
* [ISO10646及UNICODE 漢字集碼](http://www.cns11643.gov.tw/AIDB/encodings.do#encode9)
* [GB 2312](http://www.cns11643.gov.tw/AIDB/encodings.do#encode10)
* [GBK](http://www.cns11643.gov.tw/AIDB/encodings.do#encode11)
* [電信碼](http://www.cns11643.gov.tw/AIDB/encodings.do#encode12)

#### 【 CNS11643 國家標準中文交換碼 】

##### 壹、編訂緣起

民國69年9月行政院國家科學委員會集合國內編碼專家、學者在溪頭開會，達成初步原則，並據此報請行政院核定國家中文資訊標準交換碼編碼原則，翌年9月2日行政院函令國科會依據已選定之原則，邀集教育部、中央標準局及行政院主計處電子處理資料中心組成專案作業小組，積極推動編碼工作。71年7月曾編定常用字碼一種，但所收字數不夠；72年5月9日行政院資訊推動小組再次確立編碼方式，即於5月12日組成編碼技術作業小組，針對已定之編碼原則，進行編碼細則之研討，10月底完成「通用漢字標準交換碼」，並決議試用二年。試用期滿後，國科會與院主計處於74年8月邀集各相關單位與業者組成技術小組，檢討試用結果、修訂編碼原則後重編，75年3月獲行政院核定，正式公布實施。75年8月獲中央標準局審定頒布為國家標準，編號「CNS11643」；81年該局再因應各界之需要，由原２個字面 (13,051字) 大幅擴編為７個字面(48,027字)，5月公布並更名為「中文標準交換碼 (Chinese Standard Interchange Code)」。

##### 貳、適用範圍：

本標準適用於中文資訊之處理。

##### 參、編碼之考慮

|  |  |
| --- | --- |
| 一、 | 以教育部所公布的四個字體表之字集為範圍。 |
| 二、 | 根據使用的頻率及範圍，整理後分別編排於各個字面，以適應各個層次之使用者。 |
| 三、 | 符合國際資訊傳輸上所使用之CNS 5205「資訊處理及交換用七數元碼字元集(組)」及CNS 7654「資訊處理--七位元及八位元碼字元集--延碼技術」標準通信定則。 |
| 四、 | 涵蓋常用之外語字母及工商界與學校所使用之文字及符號。 |

##### 肆、字集編排原則

|  |  |
| --- | --- |
| 一、 | 「中文標準交換碼」共分為十六個字面，每個字面可陳列94列 \*94行，即8,836個字符。第一至第十一字面為標準區，第十二至第十六字面則為使用者加字區，供使用者暫編未收於標準區之字符。 |
| 二、 | 各字面字集之排列，大抵以使用頻率為次序。第一字面以常用字為主，第二字面以次常用字為主，第三字面以部分罕用字及較常用異體字為主，第四字面以ISO DIS 10646第二版之漢字、各單位╱資訊業用字及戶政用字為主，第五字面以罕用字為主，第六、七字面以異體字為主；其中第一、第二字面字集先於75年8月4日公布為國家標準。 |

##### 伍、字碼編排原則

|  |  |
| --- | --- |
| 一、 | 文字之選擇及字體悉依教育部「國字標準字體表」為基準。  說明： 中國文字的困擾主要有兩方面，一是文字的數量太多，二是異體字的增加。目前中文的總數繁多，實際上一般人常用的不過五千字左右，新字又不斷的增加，造成中文資料處理上的困擾；而教育部的標準字體表之字集係經多年之蒐集、考證、分析、選取，為較不偏頗，最具客觀性之用字字集，應能符合一般使用者之需求。 |
| 二、 | 以2個位元組 (byte)為中文碼編碼單位，並以十六進位制之文數字表示  說明： 「以2個位元組為字碼單位，於處理時可提高效率,且傳輸時可增加資訊傳輸之速度」，符合一般資料處理作業之需要。採用十六進位制之文數字編碼，係因應資料處理人員所慣用之進位法，用以表示兩位元組最為簡明。 |
| 三、 | 符合CNS 5205及CNS 7654之通信定則。  說明： 本編碼為符合CNS5205及CNS7654通信定則之規定，所有控制碼均予避開，即字碼中之00至20以及7F均予避開，則7bit字碼集共有94個編碼位置，兩個位元組共可編8,836個字碼，訂為一字面。 |
| 四、 | 依字之使用頻率編排於各字面。  說明： 在做資訊傳輸時，若欲傳送出現在不同字面上的字，必須先送出轉字面控制碼。為提高傳輸效率，將常一起出現的字編在同一字面中，可減低字面轉換的次數。 |
| 五、 | 依先筆畫後部首排列順序編碼。  說明： 每一字面均以文字總筆畫數為首序、部首為次序、筆順為末序編訂字碼，使用者只需以書寫之實際筆畫數即可查尋到國標碼。 |

##### 陸、字集說明

|  |  |
| --- | --- |
| 一、 | **標準區**  第一字面：為減少字面轉換次數，將最常用之中文字及符號、字母、部首等編於第一字面；所編字彙及碼區如下：  1. 符號區  符號區之編碼位置規劃於第一字面之2121至427E，共有3,102個編碼位置，目前暫編符號684個如下：  (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,有阿拉伯數字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)中文部首213個。  (12)控制碼符號33個。  2. 中文字區  **第一字面：**  字碼區間4421至7D4B，共編入中文字5,401個；字集來源除教育部「常用國字標準字體表」所列4,808字外，並優先收編國中、國小教科書中常用字587字及異體字6字。  **第二字面：**  字碼區間由2121至7244，故編入中文字7,650個；字集來源除教育部「次常用國字標準字體表」所列6,330字外，並篩選編入教育部「罕用國字標準字體表」表中使用頻率較高之1,320字。  **第三字面：**  字碼區間2121至6246，共編入中文字6,148個；字集來源為77年6月行政院主計處電子處理資料中心暫編之使用者加字區第14字面前段。  **第四字面：**  字碼區間2121至6E5C，共編入中文字7,298個；字集來源除77年6月行政院主計處電子處理資料中心暫編之使用者加字區第14字面後段171字外，並加入戶役政及其他使用單位、ISO 10646第2版漢字集、資訊業次常用字7,127字。  **第五字面：**  字碼區間2121至7C51，目前編入中文字8,603個字；字集來源為未編入前4個字面之教育部罕用字。  **第六字面：**  字碼區間2121至647A，共編入中文字6,388個；字集來源為未編入前5個字面，且筆畫在14畫(含)以下之教育部異體字。  **第七字面：**  字碼區間2121至6655，故編入中文字6,539個；字集來源為未編入前6個字面，且筆畫在15畫(含)以上之教育部異體字。 |
| 二、 | 使用者加字區  為適應各種不同性質之中文資料處理作業，CNS11643特別預留第十二字面至十五字碼為使用者加字區；尚未收編於本國家標準之中文字或符號，由使用視需要先編於此區內使用。  CNS11643修訂擴編後，編碼之中文字數達48,027個，已涵蓋教育部所頒訂之常用、次常用、罕用及異體四個「國字標準字體表」所有文字；惟自我國戶役政單位全面實施電腦連線作業後，其建置之全國人民基本資料庫用字，在國家標準之範圍外又增加約三萬餘姓名用字；為使此類字碼得以進行資料傳輸與交換，行政院主計處電子處理資料中心暫編其交換碼於使用者加字區：  第十五字面：字碼區間2121至6D39，共編入中文字6,831個；字集來源為戶役政作業第十五字面。我國戶役政作業使用EUC碼，其編碼原則與 CNS11643相同，為使各界易於瞭解，沿用其字集與字面之定義。惟戶役政第15字面原編7,169個字中，包括2個自重字及336個與國標碼前七字面重複字；為避免一字二碼，重複部份予以刪除;又為避免增加戶役政單位繁複的轉碼過程，重複字刪除後，原編碼位置保留空位不編碼。 |

##### 柒、CNS11643之使用

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 一、 | 字面之指定  由 1B 24 29 [F] 四個位元組之逸出順序碼指定於 G1 字元集、 1B 24 2A [F] 指定於 G2 字元集、 1B 24 2B [F] 指定於 G3 字元集，其中終結字元 [F] 可用 30～3F 來指定相對之一至十六中文字面，另十七至八十字面僅能以 1B 24 2B [I] [F] 指定於 G3字元集，其中[I][F]自 21 30 ~ 21 3F 為十七至三十二字面、22 30 ~ 22 3F 為三十三至四十八字面、23 30 ~ 23 3F 為四十九至六十四字面、24 30 ~ 24 3F 為六十五至八十字面；至於英文之字集(ASCII)則可經由 1B 28 42 指定於G0字面。在7個位元的環境下。 目前 CNS11643 的1~7 字面之終結字元 [F] 已獲國際標準組織 ISO 正式登記為 47 至 4D ，亦可使用於字集之指定。CNS11643 各字面原指定之終結字元與 ISO 之終結字元對應如下：   |  |  |  | | --- | --- | --- | | CNS字面 | CNS之終結字元 | ISO之終結字元 | | 第一字面 | 30 | 47 | | 第二字面 | 31 | 48 | | 第三字面 | 32 | 49 | | 第四字面 | 33 | 4A | | 第五字面 | 34 | 4B | | 第六字面 | 35 | 4C | | 第七字面 | 36 | 4D | |
| 二、 | 字面之轉換  1. 利用SI使用G0字面，並為鎖定方式。  2. 利用SO使用G1字面，並為鎖定方式。  3. 利用LS2使用G2字面，並為鎖定方式。  4. 利用LS3使用G3字面，並為鎖定方式。  5. 利用SS2使用G2字面，並為非鎖定方式。  6. 利用SS3使用G3字面，並為非鎖定方式。  其中 SI、SO、LS2、LS3、SS2、SS3 為調用控制符接逸出順序後以示字面調用；為求使用方便，終端設備在開機時可將G0、G1、G2等三個字集分別設定為ASCII、第一字面及第二字面,將G3字集設定為其他較常用的字面。  ＊註:關於以上控制碼的詳細使用，請參考CNS 7654 。 |

##### 捌、CNS11643之推廣應用

CNS11643碼系統依國家標準法之規定，應由經濟部標準檢驗局(由原中央標準局裁撤後標準業務歸併)負責檢討增修，為加強推廣該標準之應用，該局特將此系統及中文字型檔委託行政院主計處電子處理資料中心代為辦理推廣。該中心為維持本系統中文字型檔之完整性，以利此項國家標準之推廣，另再商得內政部及經濟部工業局同意一併對各界提供免費使用其製作之字型檔。CNS11643目前之應用情形如下：

|  |  |
| --- | --- |
| 一、 | 國內之應用  1. CNS11643 已納入「政府機關資訊處理共通規範」中，是大多數國內外資訊廠商共同遵循的中文作業系統參考準則。  2. 政府機關公文電子交換之標準傳遞碼：本院政府機關公文電子傳遞作業已規定，凡經「交換中心」(設於交通部管資中心)傳送之公文，一律先轉換為CNS11643碼。  3. 國內大型資訊系統之應用：最具代表性也是最重要的就是全國戶役政系統。目前戶役政全國連線系統是建構在MITUX系統上，屬於主從模式架構，內碼採用 UNIX系統之EUC碼。EUC碼雖與CNS11643長度不同，但卻採用了CNS11643之編碼架構及字集，因此亦可視為是CNS11643應用於內碼之實例。  4、行政院研考會所推動之BIG-5E字集(BIG-5碼擴編部分)，亦以CNS 11643為藍本，共納編CNS第1字面之3個部首字及第3字面之3,891個、第4字面之59個一般文書上常用中文字。  5. 國內外資訊廠商大都已提供其內碼與CNS 11643間的互換公式，以及叫用工具，可協助用戶進行中文資訊的交換。 |
| 二、 | 國外之應用  ISO10646 及UNICODE目前共收編26,783個漢字，其中22,892個係納入CNS 11643第1、2字面及第3字面之6,073字、第4字面之2,975字、第5字面之395字、第6字面之196字、第7字面之133字及第15字面之 86字；我國的國家標準得與國際標準相容，不但能提升國內電腦業者在國際市場的競爭力，將來ISO10646及UNICODE發展成熟後，現用中文碼亦得以順利轉換。 |

#### [top](http://www.cns11643.gov.tw/AIDB/encodings.do#top)

BIG-5碼介紹

1.簡介

BIG-5碼，係由資策會於1984年策劃制定，宗旨原是儘量不使用到控制碼範圍，並配合國人自制的五大(BIG-5)套裝軟體。由於委託民間設計，導致初期的BIG-5碼並不能使用五大套裝軟體。雖然如此，市面上絕大多數的套裝軟體都是在BIG-5內碼系統發展出來的，因此目前市面上有2-3個BIG-5碼版本，對使用者來說很難明白其中差異，所以在2003年由財團法人中文數位化技術推廣基金會接受經濟部標準檢驗局委託，召集國內業者代表、專家和學者，就BIG-5編碼字元表原始版本和各主要業界版本予以重整之最新版本，其排列規則說明如下：

2.BIG-5碼的字集

BIG-5碼系統為兩位元組之內碼系統，共可定義19782個字碼，其高、低位元組的範圍如下：

|  |  |
| --- | --- |
| 高位元組 | Ａ１Ｈ　～　ＦＥＨ　（＊１２６）  ８ＥＨ　～　Ａ０Ｈ  ８１Ｈ　～　８ＤＨ |
| 低位元組 | ４０Ｈ　～　７ＥＨ　（＊１５７）  Ａ１Ｈ　～　ＦＥＨ |

在本系統中，我們在上述的範圍內，規劃出標準字、特殊符號及使用者造字的區域，分別說明如下：

##### ☆標準字(STDFONT)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 使用範圍 | 字數 | 保留範圍 | 字數 |
| [常用字](http://www.cns11643.gov.tw/web/word/big5/stdfonta.html) | Ａ４４０～Ｃ６７Ｅ | ５４０１ | Ｃ６Ａ１～Ｃ８ＦＥ | ４０８ |
| [次常用字](http://www.cns11643.gov.tw/web/word/big5/stdfontb.html) | Ｃ９４０～Ｆ９Ｄ５ | ７６５２ | Ｆ９Ｄ６～Ｆ９ＦＥ | ４１ |
| 合　計 | １３０５３ | | ４４９ | |

※標準字中：兀(A461、C94A[刪除])與 嗀(DCD1、DDFC[刪除]) 兩個字重碼

※BIG5-ETen 與CP950中的[倚天字](http://www.cns11643.gov.tw/web/word/big5/eten.html)使用次常用字保留範圍共41字

##### ☆特殊符號(SPCFONT、SPCFSUPP)

1.各種符號區(SPCFONT)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 使用範圍 | 字數 | 保留範圍 | 字數 |
| [標準字](http://www.cns11643.gov.tw/web/word/big5/spcfonta.html) | Ａ１４０～Ａ３ＢＦ | ４０８ | －－－－－－－－－ | －－－ |
| 控制碼 | Ａ３Ｃ０～Ａ３Ｅ０ | ３３ | Ａ３Ｅ１～Ａ３ＦＥ | ３０ |
| 合　計 | ４４１ | | ３０ | |

※CP950的歐元符號(€)使用控制碼保留範圍A3E1位置

2.罕用符號區(SPCFSUPP)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 使用範圍 | 字數 | 保留範圍 | 字數 |
| 標準字 | Ｃ６Ａ１～Ｃ８ＦＥ | ４０８ | －－－－－－－－－ | －－－ |
| 合　計 | ４０８ | | ０ | |

※BIG5-ETen中的[倚天擴充字](http://www.cns11643.gov.tw/web/word/big5/kkk.htm)使用罕用符號區C6A1~C8D3範圍，內容有日文假名、俄文等特殊符號

※BIG5-2003中取消 〃(C6DE)、仝(C6DF)以及BIG5-ETen中C7F3~C8D3範圍所定義的俄文與特殊符號

##### ☆使用者造字(USRFONT)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 使用範圍 | 字數 | 保留範圍 | 字數 |
| 第一段 | ＦＡ４０～ＦＥＦＥ | ７８５ | －－－－－－－－－ | －－－ |
| 第二段 | ８Ｅ４０～Ａ０ＦＥ | ２９８３ | －－－－－－－－－ | －－－ |
| 第三段 | ８１４０～８ＤＦＥ | ２０４１ | －－－－－－－－－ | －－－ |
| 合　計 | ５８０９ | |  | |

3.各種BIG5碼的比較

台灣地區所使用的BIG5碼主要版本：

|  |  |
| --- | --- |
|  | 說明 |
| BIG5-1984 | 最早由資策會所定的版本 |
| BIG5-ETen | 倚天版本 |
| CP950 | 微軟所使用的版本 |
| BIG5-2003 | 2003年由財團法人中文數位化技術推廣基金會接受經濟部標準檢驗局委託，召集國內業者代表、專家和學者，就BIG-5編碼字元表原始版本和各主要業界版本予以重整之最新版本 |
| BIG5-IBM | IBM所使用的版本 |

##### BIG5-2003與各版本BIG5碼比較表：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 版本 | BIG5-2003 | BIG5-1984 | BIG5-ETen | Microsoft-CP950 | BIG5-IBM |
| 使用者造字區  (8140 - A0FE) | 有 | 無 | 有 | 有 | 無 |
| 符號區  (A140 - A2CE) | 有 | 有 | 有 | 有 | 有 |
| 全形英文字母  (A2CF - A343) | 有 | 有 | 有 | 有 | 有 |
| 全形希臘字母  (A344 - A373) | 有 | 有 | 有 | 有 | 有 |
| 注音符號  (A374 - A3BF) | 有 | 有 | 有 | 有 | 有 |
| 控制符號  (A3C0 - A3E0) | 有 | 有 | 無 | 無 | 有 |
| 歐元符號  (A3E1) | 有 | 有 | 無 | 有 | 無 |
| 保留  (A3E2 - A3FE) | 有 | 有 | 無 | 無 | 無 |
| 常用字  (A440 - C67E) | 有 | 有 | 有 | 有 | 有 |
| 數字符號  (C6A1 - C6BE) | 有 | 無 | 有 | 有 | 有 |
| 部首  (C6BF - C6D7) | 有 | 無 | 有 | 有 | 有 |
| 罕用符號  (C6D8 - C6E6) | 有 | 無 | 有 | 有 | 有 |
| 日文平假名  (C6E7 - C77A) | 有 | 無 | 有 | 有 | 有 |
| 日文片假名  (C77B - C7F2) | 有 | 無 | 有 | 有 | 有 |
| 保留  (C7F3 - C8FE) | 有 | 無 | 有  (C7F3-C8D3) | 有 | 有\* |
| 次常用字  (C940 - F9D5) | 有 | 無 | 有 | 有 | 有 |
| 七個倚天外字集的擴充字  (F9D6 - F9DC) | 有 | 無 | 有 | 有 | 有 |
| 表格符號  (F9DD - F9FE) | 有 | 無 | 有 | 有 | 有 |
| 使用者造字區和新常用字  (FA40 - FEFE) | 有 | 無 | 有 | 有 | 無 |

\*：僅編碼(C7F3 - C878)、(C8CD-C8D3)。

##### **BIG-5碼使用範圍表**

#### [top](http://www.cns11643.gov.tw/AIDB/encodings.do#top)

【BIG-5E 碼 】

##### 壹、編訂緣起

行政院主計處電子處理資料中心奉院交辦協助解決眾多使用BIG-5碼之政府單位，於進行公文電子傳遞時所遭遇之自造字無法轉換問題；經數次會商後，決議請 本院研究發展考核委員會成立專案委託財團法人中文數位化技術推廣基金會辦理「BIG-5碼字集擴編計畫」。86年7月擴編完成「BIG5+碼」，但此項成 果未為多數廠商採用，故使用者並無相關產品可用。 由於該擴編計畫中所完成之「標準字集」，為政府單位一般文書最常用之自造字，如應用於BIG-5碼的造字區，可整合使用者常用的自造字、降低轉碼的頻率。 有鑑於此，行政院研考會即再度委託中推會，由BIG-5E碼之「標準字集」中選取3,954個字，在BIG-5碼的造字區中建置「BIG-5碼補充字集 (BIG-5 Extension Character Set，簡稱BIG-5E字集)」；並配合行政院「電子化╱網路化政府計畫」之推動，於公文電子交換作業規範中訂為可處理中文碼類別之一。

##### 貳、BIG5+之編碼原則

(一) 長度仍為雙位元組，即高位元組之第一位元(MSB)=1。

(二) 保留原有之「標準字集字區」及「使用者加字區」架構，使與原系統具相容性。

(三) 以國家標準CNS11643及國際標準ISO10646漢字集為字源範圍，並依CNS11643之序編碼。

(四) 包含於ISO10646或CNS11643字集內且市面已廣為使用之倚天自造字及符號編入「標準字集」，並保留原碼位。

(五) 單獨成字之部首(如金、木、水、火、土)不再重覆編碼。

六) 重複字刪除其後者，錯字則依照CNS11643修正之。

##### 參、字集來源

|  |  |
| --- | --- |
| 一、 | BIG5+的字源  BIG- 5碼系統之內定字集與CNS 11643第一、二字面相同，故86年7月擴編計畫所完成之BIG5+碼亦以CNS11643為藍本，共完成「標準字集」與「推薦字集」兩部分。「標準字 集」的4,670個字符，均為一般文書常用之中文字，其中4,145個字包含於CNS11643 第3字面字集，219個字包含於第4字面字集；「推薦字集」的3,250個字符，為廠商所蒐集之使用者常用字。 |
| 二、 | BIG-5E的字源  BIG-5E碼的字源，以BIG-5+碼字集為主要來源；由於BIG-5碼造字區空間有限，BIG5+碼字集無法全部納入，故只挑選CNS11643及ISO10646漢字集所共有且為政府機關較常用的字3,954個。 |

##### 肆、字碼架構

|  |  |
| --- | --- |
| 一 | BIG+之編碼區間  BIG-5之總碼位原有19,782個，BIG5+之總碼位由19,782個擴大為23,940個(高位元組為81-FE，低位元組為40-7E、80-FE)，BIG5+之編碼字集如下：   1. 第一標準字集：此區即原BIG-5碼標準字集，但刪除2個重複字，字碼 區間A140-F9FE(高位元組為A1-F9，低位元組為40-7E、A1-FE)。共編入13,973個字符，包括常用國字5,401個(A440- C67E)、次常用國字7,693個(C940-F9D5)及符號471個(A140-A3FE)、字符408個(C6A1-C8FE)。 2. 第二標準字集：此區即擴編部份，字碼區間8180-FEA0(高位元組為81-F9，低位元組為80-A0)，共編入中文字4,158個。 3. CMEX推薦字集：因BIG-5碼系統之編碼位置有限，未能編入第一及第二標準字集之較常用罕用及異體國字、簡體字與日韓漢字3,454個，經中推會 (CMEX)建議集中收編於此區。字碼區間為原造字區之8140-83FE、8E40-A0FE(高位元組為81-83、8E-A0，低位元組為 40-7E、A1-FE)。 4. 造字區：僅使用第一、二標準字集時，可供使用者造字碼位仍為5,809個；但同時使用推薦字集時，因BIG5+碼之推薦字集係使用原造字區之 8140-83FE及8E40-A0FE，供使用者造字之碼位僅餘2,355個，可編碼區間為FA40-FEFE(785個碼位)、 8440-8DFE(1,570個碼位)。 |
| 二 | BIG-5E之編碼區間   1. 8E40 - 8E42：納編CNS11643第1字面的3個部首(原倚天定義之C6C2、C6C5、C6C6)。 2. 8E43 - A0FE：納編CNS1643第3字面的2,980個中文字。 3. 8140 - 86DF：納編CNS11643第3字面的911個中文字。 4. 86E0 - 875B：納編CNS11643第4字面的59個中文字。 5. 875C - 875C：國字零(Ｏ)。 6. 875D - 87EE：保留碼位128個。 BIG-5E碼之安裝工具與資料檔，已建置於研考會網站中，各界人士均可由該會資訊網(網址：[www.rdec.gov.tw/big-5e/bigindex1.jsp](http://www.rdec.gov.tw/big-5e/bigindex1.jsp)取用。 |

#### [top](http://www.cns11643.gov.tw/AIDB/encodings.do#top)

【 Big-5碼香港增補字符集(zh\_HK.hkbig5) 】

提供系統層級的支援，以支援香港地區big5增補字符集(Hong Kong Supplementary Character Set，HKSCS)。HKSCS集合香港特別行政區(Hong Kong Special Administration Region，簡稱HKSAR)於1999年9月所訂定的4,702個字元，專用於香港地區，作為香港的電腦計算要求通用字元集。

#### [top](http://www.cns11643.gov.tw/AIDB/encodings.do#top)

【 EUC 碼 (Extend UNIX CODE) 】

##### 壹、字碼架構

* 為UNIX作業系統使用之內碼(Extend Unix Code，EUC)
* 字碼長度：4 BYTE

##### 貳、應用現況

全國戶役政單位使用：EUC碼長度4 BYTE，故能納編高達七、八萬的姓名用字，滿足全國人民戶籍登記作業需求。

##### 參、與CNS11643的關係

* 長度雖與CNS11643碼不同，但卻採用了CNS11643之編碼架構及所有字集。
* 與CNS11643之對應：如CNS 12121 = EUC 8EA1A1A1h(h表16進位)
  1. 1st byte : 8Eh (固定值)。
  2. 2nd byte : A0h + CNS字面(如：第三字面為A3h)。
  3. 3rd byte : 80h + CNS high byte。
  4. 4th byte : 80h + CNS low byte。
* EUC碼轉換CNS11643不需以對照表逐筆比對，只需取得EUC碼2nd byte中的CNS字面數(如A3h的3)，再將EUC碼3rd byte及4th byte的High Bit改設為"0"(Off)，取消其中、英文識別碼。

#### [top](http://www.cns11643.gov.tw/AIDB/encodings.do#top)

【 CCCII 碼 】

##### 壹、編訂緣起

民國六十八年，因美國急需使用電腦處理東亞語文資料，故在加州史坦佛大學召開了一個籌劃東亞圖書館自動化的會議，希望訂定中文交換碼標準作為自動化之根據。 我國那時尚無合適可用的碼，只有日本代表提出他們的國家標準JIS C6226；因此，美方在沒有其他標準的情況下，就有採用日本標準的想法。由於日本漢字的數目和字型和中文的相差甚多，實在不足以代表中國文字，並且此舉 也深遠影響到我中華文化在電腦時代的生存問題，所以我國代表和華裔美國東亞圖書館代表都強力反對。經激烈辯論後，暫時否定了日本與美方之提案，同時我方代 表亦承諾次年三月亞洲研究學會年會中，提出我們編訂的中文訊交換碼，俾與日本字碼作一比較。

我方代表返國之後，上書政務委員李國鼎及國科會、中美會等單位，集合國內一批文字學家、圖書館學家及電腦學者，組成「國字整理小組」，立即開始整理我國文 字，並解決電腦處理中文資訊遇到的技術問題。「國字整理小組」由謝清俊教授主持，張仲陶教授襄輔；其他參與工作者有王振鵠、張鼎鍾、周駿富、潘重規、周 何、楊建樵、黃克東等教授。

我國於次屆亞洲學會年會上，提出共4,808字之「中文資訊交換碼」；「中文資訊交換碼」的架構為美方接受，但要求擴大編碼字集。「國字整理小組」於七十 年完成第二批，包括17,032個正體字、11,517個異體字(詳七十一年出版之第二冊二版，七十四年出版之第二冊三版)；七十六年再發表第三批，包括 20,583正體字。前後二次除擴編53,940個字碼外，並完成64×64，32×32機讀字型；此外，為了方便電腦上的文字處理，又編製了「中國文字 資料庫」(Chinese Character Database，簡稱CCDB)，其中列出每個字屬性如部首、筆畫、讀音以及各種對應和輸入碼。

##### 貳、編碼原則

|  |  |
| --- | --- |
| 一 | 收容的字須經文字學者認定。 |
| 二 | 依部首、筆畫序編碼。 |
| 三 | 將中文的正異體字關係用碼的位置表示出來：譬如大陸用的簡體字，「中文資訊交換碼」認為是一種異體字，簡體字的字碼比正體字字碼在第一位元組碼值多 6，而其餘第二、三位元組的碼值完全相同。其他的異體字也和正體字有位置關係，即異體字第一位元組碼值比正體字的第一位元組碼值多6的倍數。這是因為正體 字佔6字面(一字面有94×94 碼位)，而異體字則放在以後的各字面，並要和它對應的正體字有上述的位置關係。 |

##### 參、字碼架構

「中文資訊交換碼」是一個比較特殊的中文碼，它用三個位元組(byte)表達一個中文字，每個位元組只用94個碼位，因此它共有830,534個編碼空間。

##### 肆、應用現況

國內外圖書館使用。

#### [top](http://www.cns11643.gov.tw/AIDB/encodings.do#top)

【 ISO10646 及 UNICODE 漢字集碼】

由於各國文字在字數、字型、使用方式和文化上之差異，自人類開始用電腦進行資訊處理以來，就面臨了不同國家間資料交換的困擾。國際間雖已有ISO 646、ISO 2022等編碼規範，可供各國據以制定其內碼或國家標準交換碼，並互相交換；但正如前面所提，不定長度的ESC序列控制碼，增加了資料處理的困難度。再說，任何國家均不可能也不必要將所有其它文字全數編碼在自己的字集標準中，而以今日變化劇烈的網際網路新世紀，要如何處理本國以外的電子訊息呢？

##### 1.ISO 10646 研訂過程

為了解決不同國家間電腦字元資訊交換的困難，自1984年起，全球陸續出現兩個組織，希望能發展出全球可以共用的編碼字集。其中之一是國際標準組織ISO與國際電工聯盟IEC合組的第一聯合技術委員會下的SC2/WG2工作小組(ISO/IEC JTC1/SC2/WG2)。該小組草擬提出的ISO 10646標準草案，就是想集結全球通用的字符集，成為一個大聯集，以滿足各國資訊交換的需求。另外一個組織則是Unicode Consortium，也設法採用新的觀念和架構，來設計適用全球的廣用碼(Universal Code；簡稱Unicode，亦有人稱萬國碼)。這兩個組織的工作及方向雖然在一開始時是各自發展，但是最後終於殊途同歸，匯合產生目前甚受矚目的ISO 10646 /Unicode 標準，提供全球語言文字與符號之表示、傳送、交換、處理、儲存、輸入和顯示的共同編碼標準，不但避免了資源的浪費，並且真正落實了統一全球文字交換標準的理想。

ISO 10646標準從1984年ISO/IEC JTC1/SC2/WG2工作組成立，到1993年ISO 10646第一部分標準：「架構及基本多語文字面」(Architecture and Basic Multilingual Plane)印行為止，前後的發展階段長達十年，其間過程大致如下：

* 1984.4 ISO/IEC
* JTC1/SC2/WG2工作組成立
* 987.3 制訂編碼架構
* 1989.1 出版 1st DP (Draft Proposal)
* 1989.12 出版 2nd DP
* 1990.12 出版 1st DIS (Draft International Standard)
* 1991.6 1st 投票通過
* 1992.6 2nd DIS 投票通過
* 1993.5 出版 ISO 10646-1 (Part I)

在ISO 10646第一版建議書草案中，將此標準定名為Multiple Octet Coded Character Set(多八位元組編碼字元集)，並說明其編碼架構為四個八位元組，共組成128個群 (Group)，每一個群有256個字面 (Plane)，每一個字面有256×256個編碼空間，其中第32群的第32個字面為基本多語文字面(Basic Multilingual Plane；簡稱BMP)。到了1993年ISO 10646-1版本，此標準定名為Universal Multiple-Octet Coded Character Set(廣用多八位元組編碼字元集；簡稱UCS)，BMP則定義在第0群第0個字面。在使用上，假如需用的字集都在BMP中，則只要使用兩個八位元組的編碼架構即可，否則就必須使用四個八位元組的編碼架構；而且這兩種編碼架構不能混合使用，不是使用兩個八位元組，就是使用四個八位元組的固定長度編碼架構，其用意就是要避免ISO 2022不定長度的缺點。

在這種情形下，參與制定ISO 10646的代表莫不儘量爭取把自己國家使用的字集編入BMP，目的就是希望能利用BMP只用兩個八位元組的優勢，取得比四個八位元組更高的資訊處理效率。BMP既是如此重要，而其編碼位置卻有限，因此如何充份利用這些編碼空間，處理更多常用字符集，就成為標準研訂時的一個重要考慮因素。原本在BMP的設計中，除編入西方拼音文字及符號外，在漢字方面只蒐集日本、南韓及中國大陸三個地區使用的漢字字集，並沒有考慮到我國的字集，我國遂積極組團以民間團體－台北市電腦商業同業公會（TCA）的名義，參加ISO 10646的制定，以爭取將我國使用的正體字編入BMP中。

基本上我國、大陸、南韓與日本所用的都是漢字，就編碼的理論而言，本不應該分開各自編碼，造成一字多碼的現象，與ISO 10646編碼之基本原則“一字一碼”相抵觸。我國及大陸的代表因此就共同提出了HCS (Han Character Set) 的構想，這個構想同時得到美國及其它國家代表的支持，進而一步步發展至今，在1993年公佈的ISO 10646-1之BMP中，已經將台灣、大陸、日本和南韓所使用之漢字整理整合成獨立的中日韓認同表意文字區(CJK Unified Ideographs)，並予以編碼，使得亞洲幾個主要使用漢字的國家，得以有圓滿的解決方案。有關BMP字集在後文將續予介紹。

從1993年ISO 10646-1出版之後，儘管BMP　剩餘的空間十分有限，全球仍有許多國家文字等待納編入其中。而表意文字使用國家例如我國及大陸，也還有許多漢字需要擴編入此標準，因此截至目前為止，ISO 10646仍然在進行修訂與擴充。不過近幾年由於SC2/WG2與Unicode Consortium的密切合作，已經加快了ISO 10646修編的步伐，迄今已累積了近30項的標準補遺(Amendment)與技術校勘文件，十分驚人。根據WG2的預估，新版本ISO 10646可望於西元2000年出爐，而與其同步的Unicode 3.0版本可能在1999年第四季便先出版。

##### 2.ISO 10646 整體架構及BMP的中日韓表意文字字集

前面提過ISO 10646為四個八位元組的編碼架構，圖示如下：

由上圖可看出整個UCS是由128個群組(Group 00~7F)所組成，每個群組均有256個字面(Plane 00~FF)，每一字面有256個橫列(Row 00 – FF)，每一橫列是由256個單元(Cell 00–FF)所組成，所以每一字面可有256×256=65,536個碼位可定義字元。

1993年版本中只有定義上述基本架構以及BMP的字元，至於其餘群組、字面均尚無規定。但在1996年之後，由於WG2訂定UTF-16轉換格式的使用方式，對於BMP以外的字面也作了定義，大致如下：

|  |  |  |
| --- | --- | --- |
| ISO 10646之第0群組 | 編碼內容 | 編碼形式 |
| Plane 0(BMP) | 全球常用字集(含非漢字及漢字) | 雙八位元組編碼 |
| Plane 1(第一輔助字面) | 尚未編碼之非漢字字元集 | 四個八位元組編碼  使用UTF-16方式擷取字元 |
| Plane 2(第二輔助字面) | 尚未編碼之漢字字元集 |
| Plane 3 ~ 13 | 保留未來編碼用 |
| Plane 14 | Tag 文字 |
| Plane 15 ~ 16 | 用戶自定區 |

在1993年的版本中，BMP由A、I、O、R四大部份所組成。但經過增補修訂，已改為A、I、O、S、R五區，S區主要是用作UTF-16使用：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 內　容 | 碼　位 | 可編碼數 |
| Ａ區 | 拼音字母、符號和其它符號 | 0000 ～ 4DFF | 19,903 |
| Ｉ區 | 中日韓漢字區 | 4E00 ～ 9FFF | 20,992 |
| Ｏ區 | 保留未來使用 | A000 ～ D7FF | 14,336 |
| Ｓ區 | UTF-16使用區 | D800 ～ DFFF | 2,048 |
| Ｒ區 | 專用區（用戶╱業者） | E000 ～ FFFD | 8,192 |

雖然有如上的分區原則，但因編碼時的碼位及區塊的整體考慮，使得BMP現況　與上述分區有所差異。典型的例子如1998年剛擴編完成的「中日韓認同表意文字擴充A」字集6,582字，因為I區僅餘零星位置，因此就編入A區3400~4DFF的位置，而該區原為韓文拼音(Hangul)符號，但因後來韓文拼音移置O區，空出的碼位就給了「中日韓認同表意文字擴充A」，造成實際編碼與理論分區有別的現象。

在I區的中日韓漢字部份，最當初進行編碼時，因各國漢字型體不盡相同，必須先進行認同(unify)整理工作，SC2/WG2因此邀集有關各國指派專家組成CJK/JRG(中日韓聯合研究工作組，即IRG前身)，進行字集的總整理。CJK/JRG歷經五次會議完成此項艱鉅工作，所整理的「中日韓認同表意文字」(CJK Unified Ideographs)參考了我國75年版CNS 11643之第1、2、14字面(T欄)，大陸的GB 2312、GB 12345、GB 7589、GB 17590、GB 8565(G欄)，日本的JIS X 0208、JIS X 0212(J欄)及南韓的KS C 5601、KSC 5667(K欄)等標準字符集，可說已包含這四地所常用的字。其字序主要是參考康熙字典、大漢和詞典、漢語大詞典及大字源字典，以先部首後筆劃的順序排列。CJK/JRG將此結果送交SC2/WG2編碼，完成了ISO 10646:1993 之BMP中I區的表意文字編碼標準，總計含20,902個漢字。其中包含了CNS 11643用字共有17,011個字。

在BMP「中日韓認同表意文字」中，因空間有限，未能將亞洲各國所使用的表意文字全部納入，難以滿足各國實際應用的需求，因此各國均感到BMP表意文字字集有再擴增的必要。為此ISO於1993年特在SC2/WG2之下成立IRG (Ideograph Rapporteur Group)工作組，專司BMP 「中日韓認同表意文字」區的內部/水平擴充(International / Horizontal Supplementation)及外部/垂直擴充(External / Vertical Extension)之整理工作。參與者包括中國大陸、台灣、日本、韓國、香港、Unicode Consortium、美國、越南及新加坡等。

IRG成立以來，循每年兩次、各國輪流舉辦之慣例，已舉行十二次會議，討論ISO 10646 BMP表意文字字集的擴充原則，並就各國提出之字集進行認同及整理。1998年IRG完成「中日韓認同表意文字擴充A」(CJK Unified Ideographs Extension A)字集6,582字，送交SC2/WG2通過，已預定編碼於BMP之A區的3400~4DFF，其中包含之CNS 11643用字共有5,879個。此一字集與原I區「中日韓認同表意文字」最大的差別在於它在G、T、J、K欄外，新增了屬於越南用字的V欄。

經過這兩階段漢字集的整理與擴編，BMP總計已編入27,484個表意文字，幾乎已無剩餘的大塊連續編碼空間。然而漢字為數眾多，且ISO 10646的第二輔助字面(Plane 2)原本即定義專供漢字擴充用，因此ISO 10646表意文字的擴編目前仍持續進行中。現階段IRG最重要的工作是「中日韓認同表意文字擴充B」(CJK Unified Ideographs Extension B)的整理。這套字集將涵蓋康熙字典全部用字、漢語大字典全部用字及大陸、台灣、韓國、越南、香港所提的國家標準字集，其目標是使ISO 10646標準能納入更多漢字，以滿足各國使用大字集的需求。作法上則是以康熙字典、漢語大字典為基礎，挑除已編入ISO 10646之字，再加上各國所提字集，依認同規則進行整理而成。擴充B字集預定將編入ISO 10646 第二輔助字面，目前整理的總字數已達4萬多字，其中含我國CNS 11643 第4~7字面共約3萬字。預計這部份擴編工作完成後，我國CNS 11643第1至7字面字集將有95%以上的用字全部納入ISO 10646標準中，剩餘的5%是被相互認同(unified)的字。

本文主要介紹中文有關的編碼標準，對於ISO 10646 BMP其他文字的編碼無法詳加介紹，讀者可參考文末附圖，以瞭解BMP各國文字的編碼情形。以下為ISO 10646 與亞洲表意文字使用國家及我國CNS國家標準有關部份：

##### 4.UCS的編碼表示形式及使用方式

ISO 10646每一個字符都是以十六位元或三十二位元為一編碼單位，而非傳統的英文單位元組(1-byte)方式。從0000至FFFF(共65,536個碼位)，除了0000-001F(C0) 與007F-00A0(C1)等控制碼外，均可用來組成文字碼位。它的編碼表示形式共有UCS-2及UCS-4兩種：

* UCS-2：雙八位元組基本多語文字面形式(Two-octet BMP form)  
  專門用來表示BMP上的各個文字或符號，它只用到橫列和單元。
* UCS-4：四個八位元組正準形式(Four-octet canonical form)  
  用以表示ISO 10646的所有文字， 每一文字均以32位元編碼(群組、字面、橫列和單元，每個各占八位元)來表示。

雖然四個八位元組的編碼形式，使ISO 10646得以有非常大的編碼空間，但是因為對現有資訊系統而言，實施四個八位元組的編碼系統需要作較大的改變，西方國家比較不傾向使用。ISO 10646因此又定義了UTF-16轉換格式(00群組16個字面的轉換格式)，利用BMP的S區2,048個碼位，來擴充使用第00群組的第1到第16字面。其方式則是將BMP的D800到DFFF碼位分成兩部份，上半區D800～DBFF共1,024個碼位，下半區DC00~DFFF也是1,024個碼位，共可交叉定義出1,024×1,024＝1,048,576個碼位，恰可對應到BMP之後第1~16字面的每個編碼位置。如此則保留雙八位元組的優點，同時亦可使用四個八位元組的大編碼空間。這種方式在Unicode標準中的定義叫作Surrogate，只不過Unicode又區隔出一段131,072個碼位的專用區(Private used Area)，以下是兩者的比較：

|  |  |  |
| --- | --- | --- |
| 碼位區間 | ISO 10646 之UTF-16 | Unicode 2.0之Surrogate |
| D800 ~ DB7F | High-half zone | High surrogate |
| DB80 ~ DBFF | Private used high surrogate |  |
| DC00 ~ DFFF | Low-half zone | Low surrogate |

其中DB80 ~ DBFF 共128個碼位×1024=131,072就是Surrogate指定的專用區。

另外，為了符合ISO 2022的通信及傳輸規定，避開八位元碼中的控制集，ISO 10646也定義了UTF-8轉換格式，可用於本文資料的傳輸，以通過對ISO 2022八位元組結構及CNS 7656控制字元敏感的通訊系統。

##### 5.ISO 10646 的未來發展

由於國際資訊交流的日趨密切和多國語言交換之需求，已有越來越多的資訊廠商已經或正在以ISO 10646標準提供更完整之資訊處理環境。美國幾家大公司所組成的Unicode Consortium亦採納ISO 10646之BMP內容為Unicode標準，並積極進行系統的實作；中國大陸的中文訊息界專家亦在大陸積極推動ISO 10646之實作，因此ISO 10646對於中文碼標準的影響是不容忽視的。值此海峽兩岸互動關係正在改變，兩岸交流日益頻繁的時期，ISO 10646倒不失為兩岸中文碼共通標準的方案之一。海峽兩岸的中文資訊專家應可在目前ISO 10646的標準上，共同討論怎樣使ISO 10646之標準更適合兩岸之中文資訊處理，以及從資訊處理技術角度來看，如何加強雙方在這方面的合作，使得海峽兩岸的中文資訊處理技術更臻成熟。

附圖.　ISO 10646之基本多語文字面(BMP)概觀

[top](http://www.cns11643.gov.tw/AIDB/encodings.do#top)

#### 【GB2312】

GB 2312 碼是中華人民共和國國家標準漢字信息交換用編碼，全稱《信息交換用漢字編碼字符集基本集》，標準號為 GB 2312－80（GB 是“國標”二字的漢語拼音縮寫），由中華人民共和國國家標準總局發佈，1981年5月1日實施。

[top](http://www.cns11643.gov.tw/AIDB/encodings.do#top)

#### 【GBK】

GBK編碼(Chinese Internal CodeSpecification)是中華人民共和國制訂的、等同於UCS中新的中文編碼擴展國家標準。GBK編碼能夠用來同時表示正體字和簡體字，而GB2312只能表示簡體字。GBK工作小組於1995年10月，同年12月完成GBK規範。該編碼標準相容GB2312，共收錄漢字70205個、符號883個，並提供1894個造字碼位。

[top](http://www.cns11643.gov.tw/AIDB/encodings.do#top)

#### 【電信碼 】

電信碼是由交通部數據通信所編訂，全名為數據通信中文電碼，簡稱電信碼。它是一個2-byte內碼系統，可定義17,672個字碼，分二個字面。第一字面為7,902個字加41個繁體中文注音符號和94個標準 ASCII字符。第二字面為8,489個字。

[top](http://www.cns11643.gov.tw/AIDB/encodings.do#top)

﻿

全字庫服務專線：(02)2316-5300轉6822 <國家發展委員會版權所有 CopyRight 2014 ©> [cns130]