

# 「中文字構形索引序列」標準草案

---

莊德明2010.03.11

中央研究院資訊科學研究所

Email: [derming@gate.sinica.edu.tw](mailto:derming@gate.sinica.edu.tw)

URL: <http://www.sinica.edu.tw/~cdp>

---

# 中文標準交換碼

---

- ✧ CNS11643 中文標準交換碼
- ✧ CNS11643 中文標準交換碼全字庫
- ✧ CNS11643-2 中文字基礎部件及部件屬性
- ✧ CNS11643-3 中文字筆畫分類



# CNS11643

## 中文標準交換碼



# CNS11643 中文標準交換碼

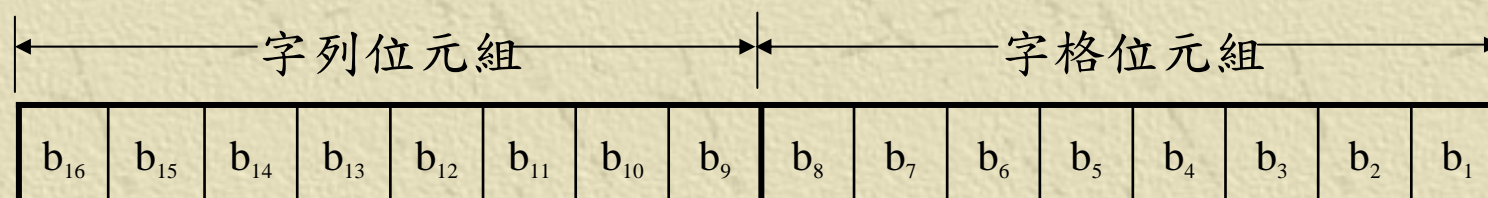
---

- ✦ CNS11643的編碼工作，始於民國69年間，歷經多次轉折，直到75年才由中央標準局公佈，取名為「通用漢字標準交換碼」（簡稱通用碼）。
- ✦ 通用碼原只收錄13,051個中文字（第1及第2字面），但主計處電子處理資料中心於民國77年另外公佈了6,148個使用者加字及編碼（於81年版編入第3字面）。
- ✦ 民國79年，中央標準局委託資訊工業策進會首度擴編通用碼為48,027個中文字（第1至第7字面），於民國81年公佈，同時更名為「中文標準交換碼」（CSIC, Chinese Standard Interchange Code）。
- ✦ 民國92年初，標準檢驗局委託財團法人中文數位化技術推廣基金會再度擴編CNS11643。

參考：<http://www.cns11643.gov.tw/cns11643/file/cns11643expand.ppt>

# CNS11643的編碼架構

- ✦ 81年版CNS11643為16字面，  
93年版CNS11643擴增至80個字面。
- ✦ CSIC的正規編碼為兩個位元組：



依ISO2022的規定，每一位元組的值為  
33~126（16進制21~7E），故每一字面為  
94X94=8,836個字格（編碼位置）。



CNS11643

中文標準交換碼全字庫



# 全字庫建置目的

---

- ✦ 建設我國的中文電腦應用環境。
- ✦ 解決個人電腦中文字數不足問題
- ✦ 解決自造字交換問題
- ✦ 解決機關、企業、團體內部同字不同碼的問題
- ✦ 協助機關、企業、團體整合及管理個人電腦上的中文字集
- ✦ 解決網頁上自造字顯示問題

參考：<http://www.cns11643.gov.tw/AIDB/intro.do>

# 全字庫系統功能

---

- ✦ 中文碼查詢：5.0版全字庫目前可供查詢的字共87,047個中文字。
- ✦ 字型下載
- ✦ 中文碼轉換
- ✦ 共用(相同)造字集安裝
- ✦ 機關企業團體自造字集整合及管理
- ✦ 網頁上自造字顯示
- ✦ 內部網路複製全字庫
- ✦ 安裝BIG-5E字集
- ✦ 轉碼閘門
- ✦ 新增字申請





CNS11643-2

中文字基礎部件及部件屬性



# 中文字基礎部件及部件屬性

---

- ✿ 部件：由筆畫組成具有組配中文字功能之構字單位。例如「言」、「牛」、「射」、「相」等皆為構成中文字之部件。
- ✿ 基礎部件：依CNS11643-2規定不得再分析成其他部件者，作為中文字結構分析最底層。例如，組合成「射」字部件「身」和「寸」，以及組合成「寶」字部件「宀」、「王」、「缶」、「貝」，六者皆是基礎部件。

參考：CNS11643-2中文字基礎部件及部件屬性

# 中文字基礎部件及部件屬性(續)

---


- ✦ 主部件：部件主要字形或寫法。例如「水」、「火」、「糸」為主部件。
- ✦ 附部件：主部件小幅度變化字形或寫法。例如「汜」、「灬」、「紕」為附部件。
- ✦ 中文字基礎部件表係以CNS11643第1及第2字面13,051中文字逐一進行分析，歸納與統計後所制定，總計385組，517個基礎部件。

## 中文字基礎部件及部件屬性(續)

---

✦ 中文字部件屬性表係將CNS11643第1及第2字面的13,051個中文字逐一分析成由基礎部件組成之部件屬性。

✦ 例如「謝」字的部件屬性為「言身寸」、「福」字的部件屬性為「礻一口田」、「瀝」字的部件屬性為「氵厂禾禾止」。



CNS11643-3  
中文字筆畫分類



# 中文字筆畫分類

---

- ✦ 筆畫：構成中文字字形之最小單位
- ✦ 筆形：筆畫之具體形狀，其主要特徵包括書寫方向及形狀。
- ✦ 筆順：由中文字之組成筆畫依其書寫順序排列而成筆畫或筆形序列。
- ✦ 筆畫分類：筆畫依其筆形主要特徵可區分成橫、豎、撇、點、折五大類。

參考：CNS11643-3 中文字筆畫分類

## 中文字筆畫分類(續)

---

- ✦ 主筆形：五大類筆畫之代表性筆形。
- ✦ 附筆形：衍生自某一主筆形或依其主要特徵可視為與某一主筆形同屬一類之其他筆形。
- ✦ 筆畫併類表計收錄筆形為「橫」者2個、「豎」者1個、「撇」者2個、「點」者4個、「折」者26個，合計35個。

表 1 筆畫併類表

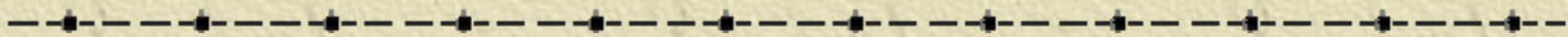
分類	主筆形		附筆形							
1	一	橫	ㄟ	橫挑						
2	丨	豎								
3	ノ	撇	ㄩ	豎撇						
4	丶	點	、	頓點	ㄨ	捺	ㄩ	挑		
				折數 1		折數 2		折數 3		
5	ㄗ	折	ㄗ	橫折	ㄥ	橫折橫	ㄣ	橫折橫折	ㄣ	橫折橫折鉤
			ㄨ	橫鉤	ㄣ	橫折彎	乙	橫折曲鉤		
			フ	橫撇	ㄩ	橫折鉤	ㄣ	豎橫折鉤		
			ㄥ	豎折	ㄣ	橫斜鉤				
			ㄣ	豎彎折右	ㄣ	豎橫折 1				
			ㄣ	豎彎折左	ㄣ	豎橫折 2				
			ㄣ	豎挑	ㄣ	豎曲鉤				
			ㄣ	豎鉤						
			ㄣ	臥鉤						
			ㄣ	斜鉤						
			ㄣ	彎鉤						
			ㄣ	撇頓點						
			ㄣ	撇橫						
			ㄣ	撇挑						
			ㄣ	挑捺						



「中文字構形索引序列」  
標準建議書



# 內容



✦ 範圍

✦ 目的

✦ 理由

# 範圍

- 
- ✦ 本標準可用來索引中文資訊系統的罕用字，以便進行罕用字檢索，同時可和中文標準交換碼相輔相成。

# 目的

---

- ✦ 中文標準交換碼CNS 11643收錄的中文字已超過八萬七千個，其中絕大多數為罕用字，而這些罕用字通常為中文電腦系統的缺字。
- ✦ 目前中文電腦的使用者在碰到缺字時，可在使用者造字區造字，再透過CNS 11643中文標準交換碼進行交換。
- ✦ 然而這些缺字置於網路上之後，使用者必須找到該字的CNS 11643字碼，才能進行檢索，這對使用者而言，相當不方便。
- ✦ 制定中文字構形索引序列，可讓使用者利用中文字的部件組合來索引罕用字，以便檢索。

# 理由

---

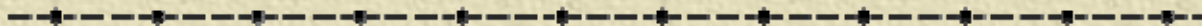
- ✿ 現行中文字碼的制定，是將中文字視同西方語言的拼音字母，而採用一字一碼，完全忽略中文字是表意文字，是由有限的部件所組成。
- ✿ 中文字構形索引序列是根據中文字的結構特性所制定，採用中文字部件的組合來索引字形，合乎中文字的特性。
- ✿ 中文字構形索引序列的設計理念雖和現有的中文字碼完全不同，但可和現有的任何中文字碼系統相容。

## 理由(續)

---

- ✦ 中研院漢籍電子文獻及數位典藏國家型計畫均採用中文字構形索引序列來解決中文字的缺字問題。
- ✦ 中文字構形索引序列係由中研院資訊所文獻處理實驗室研發，迄今已有十五年，並於今年採用公眾授權的模式釋出。
- ✦ CNS 11643目前已增訂「中文字基礎部件及部件屬性」標準(CNS 11643-2)、「中文字筆畫分類」(CNS 11643-3)，此標準可納入CNS 11643，並與CNS 11643相輔相成。

「中文字構形索引序列」  
標準草案



# 內容

---

- ✦ 適用範圍
- ✦ 用語釋義
- ✦ 符合性
- ✦ 中文字的構形索引原則
- ✦ 中文字構形索引序列
- ✦ 中文字構形索引表



## 適用範圍

- ✦ 本標準適用於索引中文資訊系統的罕用字，以便進行罕用字檢索，並可與中文標準交換碼相輔相成。

# 用語釋義－合成部件

---

✦ 合成部件：可由兩或多個基礎部件組合而成的部件。

● 例如「廂」、「霜」、「湘」等字的部件「相」，可由「木」和「目」兩個基礎部件組成。

✦ 例如「勛」、「圓」、「隕」等字的部件「員」可由「口」和「貝」兩個基礎部件組成。

● 例如「相」與「員」兩字皆為合成部件。

## 用語釋義—部件的層級

---

✿ 中文字係由一或多個部件依層次逐級組合而成，能直接體現全字造字意圖的部件稱為一級部件，一級部件以下為二級部件，以此類推。

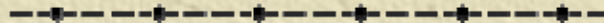
- 例如「謝」字的一級部件為「言」、「射」，其中「言」表義，「射」表音；「身」、「寸」為「謝」的二級部件。
- 例如「寶」的一級部件為「宀」、「王」、「缶」、「貝」，同時也是基礎部件，其中「宀」、「王」、「貝」表義，「缶」表音。

# 符合性

---

- ✦ 利用中文字的構形來索引字形時，若稱其部件組合規則符合本標準，則必須
  - 使用以下所規定之構字符號。
  - 依據以下所規定之構形索引序列使用原則。

# 中文字的構形索引原則



# 原則一：部件組合

---

✦ 中文字係由一或多個部件依層次逐級組合而成，絕大多數的字形都可透過各級部件的組合來索引。

- 例如「謝」字的各級部件組合「言射」、「言身寸」都可用來索引「謝」字。
- 例如「寶」字的一級部件組合「宀王缶貝」可用來索引「寶」字。

## 原則一：部件組合(續)

---

- ✦ CNS11643第1及第2字面的13,051個字中，有12,817個字都可透過各級部件的組合來索引。

## 原則二：相對位置

---

- ✿ 少數中文字的差異僅在於部件相對位置不同，此類字形的構形索引，除部件本身外，還須描述部件的相對位置。
- ✿ 此類字形通常僅有兩個一級部件，而部件的相對位置可簡化成左右、上下及內外三類。
  - 例如「暉」、「暈」二字的差別只在部件「日」、「軍」的相對位置為左右或上下；「旭」、「晷」、「晃」三字的差別只在部件「日」、「九」的相對位置為內外或上下。



## 原則二：相對位置(續)

✿ CNS11643第1及第2字面的13,051個字中，有115組、234個字的構形索引必須描述部件的相對位置，這些字如後。

- 加另叻、翌翊邺、架枷拐、旭晃昏
- 呆杏、鵠鷺、障鄣、娶媿、查杅、榷聚、聊陬、崑崙、唛媧、催催、峪峇、賀勛、吝咬、哈峇、聃呬、員唄、唯售、陪部、鄖隕、漉窪、墓墳、塵墀、幕幙、暮曛、嫫嫫、否吓、棼楸、邴陌、郇隕、吟含、岑嶺、垢壘、澆荷、阮邗、埡壘、

## 原則二：相對位置(續)

- 杓窠、邨陝、嶠晷、訛訾、坳坳、吡些、只叭、櫛  
縻、烘糞、輶輦、謨碁、邗阮、召叨、拐招、岌  
岌、可叮、邸阨、賅賚、委姝、蕩蕩、宴晏、案  
桉、哲晰、啦掬、摧摧、暈暉、主王、景暎、啼  
啻、邗防、保咻、都階、脣脰、陴郛、省眇、筭  
簪、渺涓、邾陟、晌曷、易吻、蜒蚤、幸坪、响  
响、郁隋、藹藹、郵陞、集椎、盲眈、崑岫、妃  
改、帛帕、旱盱、槻槩、旻旻、炅眈、昶咏、蜥  
蜚、杳杲、楸楸、衍洧、樞鼎、李籽、斐排、毗  
毘、泵砵、皇珀、礪碧、寫馱、麋麓、蜃蜃、蠟  
蠟、棘棗、麤鑊、蕓樞、某柑

## 原則三：重複部件

---

✦ 少數中文字是由單一部件重複組合而成，這些部件的組合方式通常為由左至右，由上至下，或呈三角狀、四角狀排列。

✦ 例如「林」字由單一部件「木」由左至右排列，「棗」字由單一部件「束」由上至下排列，「轟」字由單一部件「車」呈三角狀排列，「燚」字由單一部件「火」呈四角狀排列。

## 原則三：重複部件(續)

✪ CNS11643第1及第2字面的13,051個字中，有67個字係由單一部件重疊組合而成，這些字如下。

- 比、朋、林、珏、弱、玆、棘、兢、赫、競、姪、  
孖、屾、干、艸、田、棼、牲、砢、秝、𦍋、𦍋、  
𦍋、𦍋、𦍋、𦍋、𦍋、𦍋、𦍋、𦍋、𦍋、𦍋、𦍋、
- 爻、圭、多、炎、哥、芻、棗、爻
- 品、姦、晶、森、磊、聶、蟲、轟、轟、劦、垚、  
艸、心、弄、毛、淼、焱、焱、晶、晶、晶、  
隸、泉、香、馬、鹿、龍、龍
- 艸、焱

## 原則四：前後部件的相對位置

✿ 大多數中文字部件的相對位置，常可由前後部件的性質判斷得知。

- 例如部件「言」和其他部件組合字形時，部件的相對位置通常為左右，而「言」在左方，因此「言」、「射」組合成「謝」。
- 例如部件「雨」和其他部件組合字形時，部件的相對位置通常為上下，而「雨」在上方，因此「雨」、「相」組合成「霜」。
- 例如部件「口」和其他部件組合字形時，部件的相對位置通常為內外，而「口」在外，因此「口」、「員」組合成「圓」。

## 原則五：基礎部件不得再拆分

---

✦ 基礎部件為構形索引的最小單位，不得再拆分成其他基礎部件來索引字形。

- 例如部件「貝」為CNS 11643-2規定之基礎部件，不得再拆分成「目」、「八」，因此「目八」無法用以索引「貝」字，「𠃉王缶目八」無法用以索引「寶」字。

# 中文字構形索引序列



# 中文字構形索引序列

---





- ✦ 中文字構形索引序列可用於索引字形，它係由部件及構字符號組成。
- ✦ 構字符號分為形標及定位符號兩種，定位符號可用於描述部件的相對位置，形標則不涉及部件的相對位置。





表一、中文字部件的組合符號及組合式。

組合符號。		說 明。	部件的組合式。	
形標。		為字形標示的起點，表示結束，部件則包夾於其中。	寶 =  王 缶 貝	
定位符號。	連接符號。		謝 = 言  射。	
			霜 = 雨  相。	
			圓 = 口  員。	
	重疊符號。		二個相同部件的組合方式為由左至右。	林 =  木。
			二個相同部件的組合方式為由上至下。	棗 =  束。
			三個相同部件的組合方式為由左至右。	孖 =  子。
			三個相同部件的組合方式為由上至下。	忝 =  小。
			三個相同部件的組合方式為三角狀排列。	轟 =  車。
			四個相同部件的組合方式為由左至右。	𠄎 =    。
			四個相同部件的組合方式為由上至下。	𠄎 =  一。
	四個相同部件的組合方式為四角狀排列。	燄 =  火。		

# 構字符號

✿ 形標為字形的標示，由構字符號「形」及「」構成，其中「形」為字形標示的起點，「」表示結束，用以索引字形的部件則包夾其中。

● 例如構形索引序列「→王缶貝」可用以索引「寶」字。


✿ 定位符號分成連接符號及重疊符號兩類，連接符號主要用以描述不同部件的組合情形，重疊符號僅用以描述單一部件之重複組合。

# 連接符號

- ✦ 連接符號有三個，「 $\triangleleft$ 」為橫連，用來描述部件的組合順序為由左至右；「 $\triangle$ 」為直連，用來描述部件的組合順序為由上至下；「 $\triangleright$ 」為包含，用來描述部件的組合順序為由外至內。在構形索引序列中，連接符號應放在部件的中間。
- ✦ 例如構形索引序列「言 $\triangleleft$ 射」可用以索引「謝」字；「雨 $\triangle$ 相」可用以索引「霜」字；「口 $\triangleright$ 員」可用以索引「圓」字。

## 連接符號的併用

✦ 在構形索引序列中，不同的連接符號不可併用，而相同的連接符號可連續使用。

- 例如「寶」字的構形索引序列不可寫成「宀△王△缶△貝」，應寫成「宀△王△缶△貝」。
- 例如「謝」字的構形索引序列可寫成「言△射」或「言△身△寸」。
- 例如「晷」字的構形索引序列可寫成「日△共△言」。

# 形標內的部件順序

✦ 採用形標的構形索引序列，雖然形標內的部件順序基本上不影響字形的索引，但仍應制定部件排序之原則。目前採用之排序原則為部件間的連接順序：先假設不同的連接符號可併用，再抽離連接符號，留下的部件順序即是。

- 例如先假設「寶」字的構形索引序列可寫成「宀△王△缶△貝」，抽離連接符號後，構形索引序列應寫成「形宀王缶貝」。
- 例如假設「歡」字的構形索引序列可寫成「發△酉△欠」，抽離連接符號後，構形索引序列應寫成「形發酉欠」。

# 重疊符號

✦ 重疊符號有八個，常用的有「○○」、「○」  
「品」、「○○」四個，其中「○○」用來描述兩個相同  
部件的組合方式為由左至右，「○」用來描述兩個相  
同部件的組合方式為由上至下，「品」用來描述三個  
相同部件的組合方式為三角狀排列，「○○」用來描述  
四個相同部件的組合方式為四角狀排列。

✦ 另外四個不常用的重疊符號為「○○○」、「○○○」  
「○○○○」、「○○○○」，其中「○○○」用來描述三個相同部件  
的組合方式為由左至右，「○○○」用來描述三個相同部  
件的組合方式為由上至下，「○○○○」用來描述四個相同  
部件的組合方式為由左至右，「○○○○」用來描述四個相  
同部件的組合方式為由上至下。

## 重疊符號(續)

✦ 在構形索引序列中，重疊符號應放在部件的前面。

- 例如構形索引序列「○○木」可用以索引「林」字，「○束」可用以索引「棗」字，「品車」可用以索引「轟」字，「○○火」可用以索引「焱」字。

# 構字符號的併用

✦ 在構形索引序列中，形標和連接符號不可併用。

- 例如「謝」字的構形索引序列可寫成「言△△射」、「言△身△寸」或「形言射●」、「形言身寸●」，但不可寫成「形言△△射●」或「形言△身△寸●」。



## 構字符號的併用(續)

✦ 在構形索引序列中，重疊符號可和形標或連接符號併用。

- 例如「堯」字的構形索引序列可寫成「壺△兀」，也可寫成「品土△兀」。
- 例如「歆」字的構形索引序列可寫成「形歆西欠□」，也可寫成「形又又又又西欠□」或「形品又西欠□」。

# 一級部件的構形索引序列

✿ 由於一個字可透過各級部件的組合來索引，所以同一個字也可能有數種不同的構形索引序列，但僅有一級部件所組合的索引序列最貼近造字意圖。

● 例如「謝」字的構形索引序列可寫成「言△射」或「言△身△寸」，其中「言」、「射」為一級部件，而「身」、「寸」為二級部件，因此「言△射」最貼近「謝」字的造字意圖。

● 例如「霽」字的構形索引序列可寫成「雨△歷」、「形雨麻止□」、「形雨厂秝止□」、「形雨厂禾禾止□」，其中「雨」、「歷」為一級部件，「麻」、「止」為二級部件，「厂」、「秝」為三級部件，而「禾」為四級部件，因此「雨△歷」最貼近「霽」的造字意圖。

# 中文字構形索引表

---

✦ 中文字構形索引表係依據構形索引序列使用原則，將CNS 11643第4、5、6、7、10、11、14字面所收錄的罕用字中，共選取420個字，在兼顧構形理據及中文字碼收字的情況下，盡可能採用一級部件組合成索引序列。

# 參考文獻

---

✦ 技術報告「漢字構形資料庫的研發與應用」

- <http://cdp.sinica.edu.tw/service/documents/T090904.pdf>

✦ 缺字系統網站

- <http://char.iis.sinica.edu.tw>

✦ 中央研究院漢字部件檢字系統

- <http://www.sinica.edu.tw/~cdp/cdphanzi/>

# 感謝您的聆聽 和耐心

---

您的批評和指教  
正是我們改進的契機

---