

# 開放文學 – 漢文樂園 – 漢字基因

## 字碼 字序 字體 字形

一、字碼：原則上力求保持漢字固有形狀，於取碼之餘，使用者能記憶字形、電腦得保存漢字象徵，重點在於文化的延續，並非求一時取碼快速，滿足市場需要者。

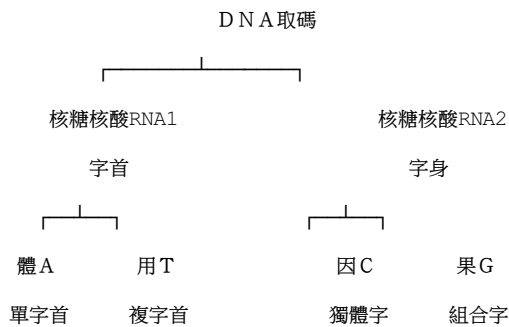
倉頡字母只有二十四個，在英文26個字母中，留出兩個供特殊發展用：x留供取碼困難時選用、z作連續字串定義之用。但因作者專心研究發展，無意市場，以致倉頡中文系統推廣失敗，使用者不多，應用者更少。

漢字分獨體與組合字，獨體可作字身；組合字由字首與字身之字碼合而成。各字皆有字碼組成漢字時，取「字首」及「字身」之碼合為一字。凡「獨體字」取一至四碼，「字首」取一至兩碼、「字身」取一至三碼；最多取五碼。

凡字首一碼已足是稱「單字首」，否則為「複字首」。

本項字碼即為「倉頡輸入法」，經台、港等地試用達三十餘年。由於早期放棄專利，推廣不易，致全部功能懸諸高閣。此字碼除供鍵盤輸入外，尚具有「字首、字身」之「理解基因」，應用於漢文自然語言，頗有成效（詳見「倉頡輸入法手冊」）。

倉頡字碼之基因結構如下：



漢字字碼種類繁多，有輸入碼、交換碼、標準碼、內碼、定義碼等。

- ①輸入碼供電腦輸入用，重點在於應用的整體規劃，否則意義不大。
- ②交換碼供系統間互換，重點在收字齊全，以便各種不同系統共用。
- ③標準碼供業界公用，但若以爛字充數，後患無窮。
- ④理解碼供程式理解之用，重點在要有理解基因，否則無法設計。
- ⑤索引碼供程式自動搜索資料用，理想之設計，可提高千百倍功效。
- ⑥內碼供各家生產廠商自用，定義自理。

為了發揮最大效益，倉頡系統將「字碼」建立在倉頡輸入法上，同時可當作輸入碼、理解碼、內碼、索引碼使用。在大量資料處理時，若要尋找某一類型、或某一筆資料，「字碼」的良窳，其效率的差異有天淵之別。

「倉頡內碼」係將24個字母各取其中5位元，五碼共25位元，壓縮為四字元後，留出7位元供各種定用。

倉頡碼在設計時，已將字形、字義分類考慮進去，以致在資料索引時，可以作正反雙向比對（以碼查概念、及以概念查碼）。此內碼為三十二位元，其中前五個位元專供文字分類，倉頡輸入碼佔二十五位元，可容約千萬字；最後二位元保留給子字集分類用。

二、字序：

倉頡輸入法有二十四個「漢字字母」供文字排序。《漢字基因字典》及「漢文史資料庫」採本字序，證明應用效率與拼音字母相等。

字序乃索引基因，在資料大量使用時，由於漢字係方塊字形，每字四字元，整齊劃一，作為索引遠較不等長之英文有效。

非從事資料處理者，不瞭解文字索引的要性，甚至有些自以為是者，不知利用字碼的特性，為了增進處理效能，另行建立序列對照表，極不直觀。每當資料建檔、搜索時，都要經過兩道或三道手續，效率損失很大。

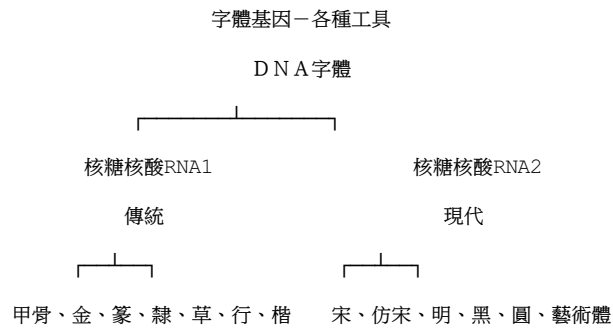
如果資料放置方式沒有直觀的順序，則完全沒有效率可言。

三、字體：

隨著環境的變遷，漢字字體因政體、書寫工具、應用方式而有所不同，傳統有：「甲骨文、金文、大篆、小篆、隸書、草書、行書、楷書」等，字形則神髓如故但筆形有別。

於客觀條件，甲骨文是刻劃在甲骨之上者，筆形銳直；金文則澆鑄於金屬器皿上者，形狀

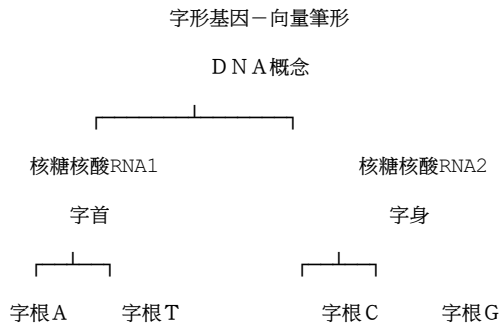
厚實；大、小篆已有毛筆寫具，字形古拙；隸書乃隸人工作所為，係用筆刷書寫；至草、行、楷毛筆盛行，故字形靈巧成熟。近代有印刷體如宋、仿宋、明、黑、圓及多種藝術體，各有特色及適應場合。



#### 四、字形（筆形）。

基本向量筆形9個，字根64個，供字庫組字用（詳見「中文字庫介紹」）。

「漢字大字庫」係根據本字形基因，以200KB系統空間，採用無級次放大，每字之組成耗中央處理器系統脈衝約300個，可組成各種字形約數千萬個。



「字形」所衍生之字庫，可以徹底解決當前中文系統文字不足與不能兼容的困境。目前完成之第六代中文字形產生器，所佔程式空間僅160KB，若用倉頡碼則不需儲存任何內碼表，有碼即有字。每個字形皆可作各種大小及字體的變化，在速度上，以3Gz CPU每秒鐘可組成16\*16之字形四萬六千個。

由於研究與商品性質不同，上述字形產生器並未推廣成為商品，記載於此係供有心人士參