



原文链接：<http://zd.hwxnet.com/article/1.html>

阅读次数：4,996



扫描二维码  
快速查阅文章

## 汉字简介

### 汉字简介

汉字是汉语书写的最基本单元，其使用最晚始于商代，历经甲骨文、大篆、小篆、隶书、楷书（草书、行书）诸般书体变化。秦始皇统一中国，李斯整理小篆，“书同文”的历史从此开始。尽管汉语方言发音差异很大，但是书写系统的统一减少了方言差异造成的交流障碍。

东汉许慎在《说文解字》中将汉字构造规律概括为“六书”：象形、指事、会意、形声、转注、假借。其中，象形、指事、会意、形声四项为造字原理，是“造字法”；而转注、假借则为用字规律，是“用字法”。

三千余年来，汉字的书写方式变化不大，使得后人得以阅读古文而不生窒碍。但近代西方文明进入东亚之后，整个汉字文化圈的各个国家纷纷掀起了学习西方的思潮，其中，放弃使用汉字是这场运动的一个重要方面。这些运动的立论以为：跟西方拼音文字相比，汉字是繁琐笨拙的。许多使用汉字国家即进行了不同程度的汉字简化，甚至还有完全拼音化的尝试。日文假名的拉丁转写方案以及汉语多种拼音方案的出现都是基于这种思想。中国大陆将汉字笔划参考行书草书加以省简，于1956年1月28日审订通过《简化字总表》，在中国及新加坡使用至今。台湾则一直使用繁体中文。

目前在使用汉语的地区，大都使用两种规范汉字，分别是繁体中文（繁体字）和简体中文（简体字）。

汉字，是记录汉语的文字系统，并仍然或曾经在日语和朝鲜语、越南语中使用。汉字是世界上最古老的文字之一，拥有4500年以上的历史。狭义地说，它是汉族的文字；广义地言，它是汉字文化圈共同的文字。

汉字是承载文化的重要工具，目前留有大量用汉字书写的典籍。不同的方言都使用汉字作为共同书写体系，因而汉字在历史上对中华文明的传播起到了重要作用，并成为东南亚文化圈形成的内在纽带。在汉字发展过程中，留下了大量诗词、对联等文化，并形成了独特的汉字书法艺术。

一个汉字一般具有多种含义，也具有很强的组词能力，且很多汉字可独立成词。这导致了汉字极高的“使用效率”，2000左右常用字即可覆盖98%以上的书面表达方式。加之汉字表意文字的特性，汉字的阅读效率很高。汉字具备比字母文字更高的信息密度，因此，平均起来，同样内容的中文表达比其他任何字母语言的文字都短。

目前的汉字体系分为繁体字和简体字，前者用于台湾、香港、澳门和北美的华人圈中，后者用于中国大陆和新加坡以及东南亚的华人社区。通常说来，两种汉字书写系统虽然有差异，常用汉字的个体差异不到25%。

由于汉字书写复杂，“汉字落后论”的说法存在了很长时间，认为汉字是教育及信息化瓶颈，并有“汉字拉丁化”甚至废除汉字的推动行为。现在一般认为汉字也有突出优点，初始学习难度虽大，但掌握常用字后不存在类似海量英文单词的继续学习问题，且其表意特性也能充分调动人脑的学习能力。在计算机输入问题基本解决后，“汉字落后论”及“汉字拉丁化”已实际上逐渐被大多数人抛弃。

目前汉字系统已经基本稳定，但汉字的规范化、生僻字的自然消亡仍在继续进行。

## 关于汉字编码

为进行信息交换，各汉字使用地区都制订了一系列汉字字符集标准。

① GB2313字符集，收入汉字6763个，符号715个，总计7478个字符，这是大陆普遍使用的简体字符集。楷体-GB2313、仿宋-GB2313、华文行楷等市面上绝大多数字体支持显示这个字符集，亦是大多数输入法所采用的字符集。市面上绝大多数所谓的繁体字体，其实采用的是GB-2313字符集简体字的编码，用字体显示为繁体字，而不是直接用GBK字符集中繁体字的编码，错误百出。

② BIG-5字符集，收入13060个繁体汉字，808个符号，总计13868个字符，目前普遍使用于台湾、香港等地区。台湾教育部标准宋体楷体等港台大多数字体支持这个字符集的显示。

③ GBK字符集，又称大字符集(GB=GuóBiāo国标，K=扩展)，包含以上两种字符集汉字，收入21003个汉字，882个符号，共计21885个字符，包括了中日韩(CJK)统一汉字20902个、扩展A集(CJK Ext-A)中的汉字52个。Windows 95\98简体中文版就带有这个GBK.txt文件。宋体、隶书、黑体、幼圆、华文中宋、华文细黑、华文楷体、标楷体(DFKai-SB)、Arial Unicode MS、MingLiU、PMingLiU等字体支持显示这个字符集。微软拼音输入法2003、全拼、紫光拼音等输入法，能够录入如镕镕烝夬喆嘉姤赉赉奕夔映堃愨晴璉等GBK简繁体汉字。

BIG-5 (繁体中文)与GB-2313 (简体中文)，编码不相兼容，字符在不同的操作系统中便产生乱码。文本文字的简体与繁体(文字及编码)之间的转换，可用BabelPad、TextPro或Convertz之类的转码软件来解决。若是程序，Windows XP操作系统，可用Microsoft AppLocale Utility 1.0解决；Windows 2000的操作系统，大概只有用：中文之星、四通利方、南极星、金山快译之类的转码软件方能解决了。

④ GB18030字符集，包含GBK字符集、CJK Ext-A 全部6582个汉字，共计27533个汉字。宋体-18030、方正楷体(FZKai-Z03)、书同文楷体(MS Song)宋体(ht\_cjk+)、香港华康标准宋体(DFSongStd)、华康香港标准楷体、CERG Chinese Font、韩国New Gulim，以及微软Windows Vista操作系统提供的宋黑楷仿宋等字体亦支持这个字符集的显示。Windows 98支持这个字符集，以下的字符集则不支持。手写输入法逍遥笔4.0版支持GB18030字符集及方正超大字符集汉字的录入。



⑤ 方正超大字符集，包含GB18030字符集、CJK Ext-B中的36862个汉字，共计64395个汉字。宋体-方正超大字符集支持这个字符集的显示。Microsoft Office XP或2003简体中文版就自带有这个字体。Windows 2000的操作系统需安装超大字符集支持包“Surrogate更新”。

⑥ ISO/IEC 10646 / Unicode字符集，这是全球可以共享的编码字符集，两者相互兼融，涵盖了世界上主要语文的字符，其中包括简繁体汉字，计有：CJK统一汉字20902个，CJK Ext-A 6582个，Ext-B 42711个，共计70195个汉字。SimSun-ExtB(宋体)、MingLiU-ExtB(细明体)能显示全部Ext-B汉字。至今尚无单独一款字体能够显示全部70195个汉字，但可用海峰五笔、新概念五笔、仓颉输入法世纪版、新版的微软新注音、仓颉输入法 6.0 版(单码功能)等输入法录入。Ext-C还有2万多个汉字。详情请参阅香港中文大学网站、马来西亚仓颉之友网站、福建陈清钰个人网站。

⑦ 汉字构形数据库2.3版，内含楷书字形60082个、小篆11100个、楚系简帛文字2627个、金文3459个、甲骨文177个、异体字12768组。可以安装该程序，亦可以解压后使用其中的字体文件，对于整理某些古代文献十分有用。

如果超出了输入法所支持的字符集，就不能录入计算机。如果没有相应字体的支持，则显示为黑框、方框或空白。如果操作系统或应用软件不支持该字符集，则显示为问号(一个或两个)。在网页上亦存在同样的情况。

### 关于Unicode

由于各国国家标准字集所收的汉字字数、常用字的差异，虽然象中国两岸GB/BIG5字集常用字基本类似，转换后阅读并不成问题，但是这种编码转换的混乱关系，对文字交流始终是一种障碍。因此相关国家的标准化组织和文字工作者经过共同努力，终于在93年完成了包含中日韩(CJK)汉字的Unicode 汉字标准ISO 10646.1。Unicode是完全双字节表示的多国文字编码体系，编码空间0x0000-0xFFFF。ISO 10646.1汉字标准使用编码0x4E00-9FA5，共包含20902个汉字。其中：大陆(S)提出的汉字17124个，台湾(T)提出的汉字17258个；S与T的并集，即中国(C)提出的汉字为20158个。日本(J)提出的汉字为12157个，中国未提出的690个(Ja)；韩国(K)提出的汉字为7477个，其中中国未提出的90个(Ka)；Ja与Ka并集共744字。支持Unicode编码的相关电脑系统软件，如Unix，Win95已有推出，但是由于Unicode的ASCII码是用双字节编码(即一般电脑系统中的单字节ASCII码前加 0x00)，同时其汉字编码与各国的现有编码也不兼容，造成现有的软件和数据不能直接使用，所以目前完全使用Unicode软件系统的用户并不多，大多数只将它此作为一个国际语言编码标准来使用。

### 精彩网址推荐

汉文学网 <http://www.hwxnet.com/>

在线词典 <http://cd.hwxnet.com/>

成语大全 <http://cy.hwxnet.com/>