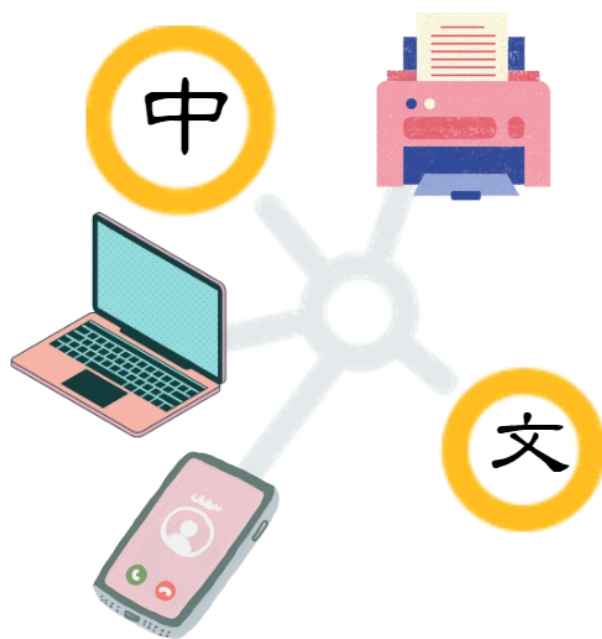


国家标准 GB 18030–2022

《信息技术 中文编码字符集》

理解与使用



编写单位：中国电子技术标准化研究院

二〇二三年四月

主编

孙文龙

副主编

范科峰 董 建 杨 磊

编写人员（排名不分先后）

黄姗姗	陈 壮	王 欣	陈晓研	孙伯君
吴 健	唐英敏	张建国	陈 恳	

顾问

王立建 代 红

版权声明

本文件版权属于中国电子技术标准化研究院。

使用说明：未经中国电子技术标准化研究院书面授权，不得以任何方式复制、抄袭、影印、翻译本文件的任何部分。凡转载或引用本文的观点、数据，请注明“来源：中国电子技术标准化研究院”。

前 言

《国家标准 GB 18030-2022<信息技术 中文编码字符集>理解与使用》是中国电子技术标准化研究院组织制定和发布的标准相关技术文件，旨在帮助读者理解强制性国家标准 GB 18030-2022 的技术内容，并提供标准化实践指引。

目录 CONTENTS

引言.....	1
第一章 标准制定概况.....	2
1.1 标准起草和颁布基本情况.....	2
1.1.1 标准修订计划.....	2
1.1.2 标准颁布和实施计划.....	2
1.1.3 标准基本内容.....	2
1.2 编制的目标和原则.....	3
1.2.1 目标.....	3
1.2.2 原则.....	3
1.3 编制过程.....	4
1.4 意见处理情况.....	4
1.5 标准发布.....	5
1.6 GB 18030-2022 国家标准第 1 号修改单情况.....	5
1.6.1 主要技术变化.....	6
1.6.2 征求意见情况.....	6
第二章 标准正文章节解读.....	7
2.1 标准规定范围和适用对象的界定.....	7
2.1.1 标准规定的范围.....	7
2.1.2 标准的适用范围.....	7
2.2 规范性引用文件.....	8
2.3 术语和定义的解读.....	10
2.4 标准收录的字汇.....	12
2.4.1 概述.....	12
2.4.2 单字节部分的字汇.....	13
2.4.3 双字节部分的字汇.....	13
2.4.4 四字节部分的字汇.....	13
2.5 标准总体编码结构.....	13
2.6 字符的排列顺序.....	16
2.7 关于码位分配的补充说明.....	17
2.7.1 CJK 统一汉字.....	17
2.7.2 CJK 兼容汉字.....	18
2.7.3 汉字部首.....	19
2.7.4 少数民族文字.....	21
2.8 部分字符和代码的说明.....	21
2.9 三个实现级别的界定.....	22
2.9.1 实现级别 1.....	23
2.9.2 实现级别 2.....	23
2.9.3 实现级别 3.....	23
第三章 标准五个附录的解读.....	25
3.1 附录 A（规范性）双字节字符表.....	25
3.2 附录 B（规范性）表意文字描述符.....	26

3.3 附录 C（规范性）四字节字符表.....	28
3.4 附录 D（资料性）部分字符和代码的说明.....	29
3.4.1 调整过的 18 个 GB/T 13000 代码位置.....	30
3.4.2 在双字节区删除的 6 个汉字构件.....	30
3.4.3 在双字节区删除的 9 个 CJK 兼容汉字.....	30
3.5 附录 E（资料性）《通用规范汉字表》汉字的代码位置.....	31
第四章 常见标准使用问题解答	32
4.1 GB 18030 标准相关问题.....	32
4.1.1 什么是编码字符集？	32
4.1.2 什么是 GB 18030 标准？	32
4.1.3 GB 18030 收录汉字的数量和范围？	32
4.1.4 GB 18030 三个版本的区别和关系？	32
4.1.5 GB 18030 与 GB/T 2312、GBK、GB/T 13000、ISO/IEC 10646、Unicode Standard 等标准和文件的关系？	33
4.2 政策相关问题.....	35
4.2.1 不符合 GB 18030 的产品如何处理？	35
4.2.2 在标准发布前已经进入流通领域的产品是否需要符合标准？	35
4.3 产品实现应关注的问题.....	35
4.3.1 GB 18030 适用于什么产品？	35
4.3.2 如何理解 GB 18030 对产品的强制性要求？	36
4.3.3 GB 18030 的 3 个实现级别分别对应什么产品？	36
4.3.4 符合 GB 18030-2005 强制部分要求的产品，是否可以认为符合 GB 18030-2022 的实现级别 1？	37
4.3.5 产品达到什么要求才算符合 GB 18030？	37
4.4 消费者使用中文信息技术产品时的常见问题.....	38
4.4.1 消费者如何判断产品是否符合 GB 18030 标准？	38
4.4.2 作为用户，是否使用了符合 GB 18030 标准的产品就能处理全部生僻字？	38
4.5 检测及认证相关问题.....	38
4.5.1 标准符合性检测和认证的定义和区别？	38
4.5.2 GB 18030 的检测和认证机构.....	39
4.5.3 什么样的产品适用于 GB 18030 标准符合性检测？	39
4.5.4 什么样的产品不适用于 GB 18030 标准符合性检测？	40
4.5.5 需要对产品中哪类功能进行标准符合性检测？	40
4.5.6 检测范围内的哪些功能应符合标准要求？	40
4.5.7 企业对产品进行自测时应注意什么？	41
4.5.8 测试中有哪些常见的检出问题？	41
第五章 标准相关参考资料选编	42
5.1 编码字符集标准的演进.....	42
5.2 汉字的排序.....	46
5.2.1 双字节 2 区汉字的分级和排序.....	44
5.2.2 双字节 3 区、4 区以及四字节区汉字的排序.....	44
5.3 中国民族古文字.....	46
5.4 CJK 统一汉字及其扩充的介绍.....	48

5.4.1 国际标准 ISO/IEC 10646 的介绍.....	48
5.4.2 CJK 统一汉字的介绍.....	48
5.4.3 汉字的认同.....	49
5.4.4 CJK 统一汉字的扩充.....	49
5.4.5 CJK 兼容汉字的介绍.....	50
5.4.6 汉字部首和汉字排序.....	52
5.5 汉字编码国际工作组（IRG）.....	54
附件.....	55

引言

编码字符集是各类软硬件产品能够处理文字信息的底层基础,服务于文字信息处理、交换、存储、传输、显现、输入和输出的各个环节。为从国家层面统一编码字符集标准,维护国家通用语言文字及少数民族语言文字在信息产业的互联互通和规范使用,我国制定了强制性国家标准 GB 18030《信息技术 中文编码字符集》。

GB 18030《信息技术 中文编码字符集》是信息技术领域少有的强制性国家标准之一,在中文信息技术标准体系中具有基石地位。该标准实现了汉字及多种我国少数民族文字的信息化处理,有力保障了中华优秀传统文化传承,支撑政务服务和公共服务行业解决姓名生僻字人群办事难题,促进语言文字信息技术产业高质量发展。

伴随着我国信息化发展战略的稳步推进,国家通用语言文字及各少数民族语言文字的信息处理与交换需求不断增长,对于编码字符数量和范围也提出了更高的要求。鉴于此,GB 18030 适时开展了修订更新工作。自 2000 年第一版 GB 18030-2000《信息技术 信息交换用汉字编码字符集 基本集的扩充》发布以来,GB 18030 进行了 2 次修订,扩充了少数民族文字收录种类与编码汉字数量。最新版的 GB 18030 已于 2022 年发布,将于 2023 年 8 月 1 日起实施。

本文件将重点针对最新版标准 GB 18030-2022 进行解读和说明,帮助读者更好地理解与使用该标准。

第一章 标准制定概况

1.1 标准起草和颁布基本情况

1.1.1 标准修订计划

国家标准化管理委员会于2017年10月11日下达2017年第三批国家标准制修订计划的通知（国标委综合[2017]103号），第1项计划编号：20171716-Q-469《信息技术 中文编码字符集》强制性国家标准修订。

1.1.2 标准颁布和实施计划

国家市场监督管理总局（国家标准化管理委员会）于2022年7月19日发布中华人民共和国国家标准公告——2022第10号，其中第1项便为GB 18030-2022《信息技术 中文编码字符集》。该标准将从2023年8月1日开始正式实施。

1.1.3 标准基本内容

强制性国家标准GB 18030-2022《信息技术 中文编码字符集》规定了信息技术用的中文图形字符及其二进制编码的十六进制表示，适用于一切同时具有中文信息处理和交换功能的技术类产品进行中文和其他文字图形字符信息的处理、交换、存储、传输、显现、输入和输出。

标准内容共包括9章和5个附录：

（1）范围；（2）规范性引用文件；（3）术语和定义；（4）字汇；（5）总体结构；（6）字符的排列顺序；（7）码位分配；（8）部分字符和代码的说明；（9）实现的级别。

附录A（规范性）：双字节字符表；附录B（规范性）：表意文字描述符；附录C（规范性）：四字节字符表；附录D（资料性）：部分字符和代码的说明；附录E（资料性）：《通用规范汉字表》汉字的代码位置。



1.2 编制的目标和原则

1.2.1 目标

——在技术标准层面解决《通用规范汉字表》的信息处理问题

2013年6月5日，国务院批准发布了《通用规范汉字表》，收录了8105个汉字。同年，教育部、工业和信息化部、国家民族事务委员会、公安部、民政部、文化部、国家工商行政管理总局、国家质量监督检验检疫总局、国家新闻出版广电总局、国家语言文字工作委员会、中国科学院中国社会科学院等十二部门联合发布《教育部等十二部门关于贯彻实施〈通用规范汉字表〉的通知（教语信[2013]2号）》，要求“社会一般应用领域的汉字使用应以《通用规范汉字表》为准”。

为在信息技术领域落实上述要求，GB 18030第二次修订（即2022版）把收录《通用规范汉字表》规定的全部汉字作为工作目标之一。

——符合2020年《强制性国家标准管理规定》的要求

2020年1月6日，国家市场监督管理总局发布了《强制性国家标准管理办法》。根据其要求，GB 18030新版不能继续采用条文强制的方式。因此，GB 18030第二次修订（即2022版）应对文本格式作相应调整，以符合该规定的要求。

本次标准编制的第二个目标是，既能符合我国信息产业发展的实际需要，又能符合《强制性国家标准管理规定》的有关要求。

——与国际标准ISO/IEC 10646同步更新中文字符

GB 18030的框架结构，自设计之初就考虑了与国际标准ISO/IEC 10646《信息技术 通用编码字符集（UCS）》进行双向转换中文字符数据的需要。GB 18030第二次修订（即2022版）的第三个目标是对国际标准ISO/IEC 10646（2017年版）中新增的中文字符进行补充收录，包括CJK统一汉字（及其扩充）和滇东北苗文等几种我国少数民族现行使用的文字。

1.2.2 原则

GB 18030第二次修订（即2022版）遵循以下原则：

——体系结构与上一版（GB 18030-2005）保持一致。

——在字符收录方面，仅收录中国各民族现行文字及其他常用文字、标点、符号如拉丁字母、阿拉伯数字等。其中：



- 保留上一版收录的全部文字、标点、符号，但其中的维吾尔文、哈萨克文、柯尔克孜文收录情况根据GB/T 21669-2008《信息技术 维吾尔文、哈萨克文、柯尔克孜文编码字符集》的规定进行修订；
- 收录2013年国务院批准发布的《通用规范汉字表》规定的全部8105个汉字；
- 根据GB/T 13000（或ISO/IEC 10646）最新版情况补充汉字和中国少数民族现行文字。

——编码与GB/T 13000（或ISO/IEC 10646）规定的编码一一对应。本标准2005版中规定的GB 13000临时码位，根据相应字符在GB 13000（或ISO/IEC 10646）最新版中的编码情况进行了调整。

1.3 编制过程

2017年12月，在接到国家标准化管理委员会下达的任务后，中国电子技术标准化研究院正式牵头成立标准修订起草组，全面启动修订工作；2018年7月，第一次公开征求意见；同年10月，完成标准送审稿审查；2019年，标准报批稿¹由工业和信息化部对外通报；2021年3月和11月，分别通过全国信息技术标准化技术委员会网站和全国标准信息公共服务平台第二次公开征求意见；2022年7月19日，标准正式发布。

1.4 意见处理情况

本标注在修订过程中充分参考各方意见建议，积极通过会议、邮件、电话等形式与有关专家进行技术内容的讨论。

在第一次征求意见阶段，标准共收到7个单位和1名个人的反馈意见共11条。意见处理情况如下：1）对于将标准名称改为“中文图形字符编码字符集”的建议，由于“中文编码字符集”一名已经包含了图形字符和非图形字符，起草组认为应保留原名称。2）对于区分列出常用汉字和非常用汉字以及简繁汉字对应关系的建议，由于《通用规范汉字表》中已有相应规定，标准在附录E给出了《通

¹ 在2020年1月6日国家市场监督管理总局发布《强制性国家标准管理办法》前，本标准第一次报批。其后，标准起草组根据《强制性国家标准管理办法》的有关规定再次修改了标准草案。



用规范汉字表》中汉字的GB 18030编码列表，便于使用者对照查询。3) 对于修正朝鲜文音节字符数量统计的建议，标准进行了相应部分的修改。

在第二次征求意见阶段，标准起草组于2020年11月邀请国家标准化管理委员会、教育部、公安部、人民银行、国家卫生健康委统计信息中心等17个单位召开了标准草案征求意见会。与会专家均同意本标准改为全文强制，并采用分级实现的方式对产品应支持的编码字符数量和范围进行要求，同时建议实现级别3的适用对象应确定为“用于政务服务和公共服务的产品”，以解决上述领域中较为突出的人名、地名生僻字问题。

此外，标准还对第二次征求意见稿全国公开征求意见期间收到的建议进行了如下处理：1) 对于跟随国际标准ISO/IEC 10646动态更新的建议，标准起草组在GB 18030国家标准第1号修改单中即补充收录了国际标准最新版本中的新增编码汉字（见本章第6节），后续也将持续关注国际标准更新情况，及时进行增补。2) 对于提供与GB/T 13000代码映射算法的建议，因少数来自GBK的汉字和部件的码位调整，无法通过已有算法得到此部分码位，因此标准中未列出具体映射算法，而是采用代码映射表文件的方式面向所有使用者公开提供（见本文件附件）。

1.5 标准发布

2022年7月28日，国家标准化管理委员会、工业和信息化部、国家语言文字工作委员会在京联合召开《信息技术 中文编码字符集》强制性国家标准发布宣贯会，受到有关部门及社会各界的广泛关注。市场监管总局党组成员、副局长、标准委主任田世宏，工业和信息化部党组成员、副部长张云明，教育部党组成员、副部长、国家语委主任田学军出席会议并讲话。国务院办公厅电子政务办公室、民政部、人力资源社会保障部、国家电子文件管理部际联席会议办公室等有关部门、标准起草单位等近50人参加了会议。

会议当日，中央电视台《新闻联播》节目播报了新版《信息技术 中文编码字符集》强制性国家标准发布的消息，新华社、光明日报等媒体也同步发布了相关新闻。

1.6 GB 18030-2022 国家标准第1号修改单情况

作为对GB 18030-2022《信息技术 中文编码字符集》的补充和修正，在工业和信息化部信息技术发展司的指导下，中国电子技术标准化研究院牵头开展了GB 18030-2022国家标准第1号修改单的研制工作。

1.6.1 主要技术变化

相比GB 18030-2022标准，修改单新增收录了：1) 公安人口信息专用字库中尚未得到国际标准或国家标准编码的汉字；2) 已经收入ISO/IEC 10646: 2020《信息技术 通用编码字符集》及其补篇但尚未收入GB 18030-2022的CJK统一汉字及其扩充，如扩充G和扩充H等。

此外，修改单还对GB 18030-2022标准中的少量编辑性错误进行了修正，如表E.1《<通用规范汉字表汉字>的字形和代码位置》中部分代码位置和字形等。

1.6.2 征求意见情况

该修改单于2022年11月至2023年2月面向社会第一次公开征求意见。基于第一次征求意见阶段收到的来自社会各界的反馈意见，标准修改单起草组进一步对技术内容修改完善，形成了第二次征求意见稿，并于2023年4月6日至5月6日公开征求意见。

相对于第一次征求意见稿，第二次征求意见稿在“公安人口信息专用字库补充汉字”部分删除了已在国际标准ISO/IEC 10646:2020及其修改单中收录的汉字、已进入审批流程并即将在国际标准下一版本中收录的汉字，和可与国际标准已编码字认同的汉字。在第二次征求意见稿中，“公安人口信息专用字库补充汉字”部分的总字数确定为614字。

第二章 标准正文章节解读

2.1 标准规定范围和适用对象的界定

2.1.1 标准规定的范围

【标准条文】

1 范围

本文件规定了信息技术用的中文⁽¹⁾图形字符⁽²⁾及其二进制编码的十六进制表示。

【条文解读】

——中文

本标准中所说的“中文”包括汉字和我国 10 种少数民族现行文字，即：蒙古文（包括传统蒙古文、满文、锡伯文、托忒文、阿礼嘎礼字）、藏文、维哈柯文（包括维吾尔文、哈萨克文、柯尔克孜文）、朝鲜文、规范彝文、滇东北苗文、傈僳文、西双版纳新傣文、西双版纳老傣文、德宏傣文。

——图形字符

图形字符是不同于控制功能的字符，通常具有书写、打印或显示的可视表示。本标准中收录的具有字形的汉字和少数民族文字都属于图形字符。

2.1.2 标准的适用范围

【标准条文】

1 范围

……

本文件适用于中文和其他文字图形字符信息的处理、交换、存储、传输、显现、输入和输出。⁽³⁾

本文件适用于具备中文和其他文字图形字符信息化处理及交换功能的技术类产品，包括但不限于以输入法、光学字符识别（OCR）、编辑校对、机器翻译、

语音合成、文字转写、智能写作等为代表的软件产品，以及以计算机、通信终端设备、电子书阅读器、学习机等为代表的硬件产品。

【条文解读】

——适用对象

本标准适用于一切同时具有中文信息处理和交换功能的技术类产品。

中文信息的交换是指采用通信手段或可交换的媒体把字符编码数据从一个用户传送到另一个用户。在中文图形字符信息交换过程中，如果收发双方采用的编码方式不同，则文字信息在传输后会出现无法正确显示的问题，即乱码。使用一套统一的国家标准编码字符集对于确保传输的文字信息在收发双方能够显示一致起着重要作用。

以下列举了一些适用本标准的代表性产品和服务：

1.具有中文信息处理功能的软件产品，如各类操作系统、字库、输入法、文字处理、语音处理、文档编辑、机器翻译、汉字简繁转换等软件产品。

注：中文输入法产品包括但不限于键盘、手写、语音输入法产品。其中，既包括作为单独商品出售的输入法产品，也包括作为商品组成部分的输入法产品，如 MS Windows、Mac OS、Linux 等桌面操作系统，以及 Android、iOS 等移动操作系统自带的中文输入软件。

2.具有中文信息处理功能的硬件产品，即安装前述软件产品的计算机、打印机、电信终端设备（如移动通信终端）、电子书阅读器、学习机等硬件产品。

3.数字出版物，如数字化教材、图书、报纸、期刊、课件、游戏、地图，以及以网络形式出版的文字作品等。

4.数字传媒，如互联网媒体，包括公共信息类网站、微信公众号等。

5.其他对语言文字进行信息化处理或内容编辑的信息技术产品，如各类应用软件。

6.具有中文信息处理需要的行业应用，如户籍、金融、交通、医疗、教育、供水、供电以及其它公共服务等。

2.2 规范性引用文件

本标准共采用了 3 项文件，它们均为国家标准，见标准条文。

【标准条文】



2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 2312—1980 信息交换用汉字编码字符集 基本集

GB/T 11383—1989 信息处理 信息交换用八位代码结构和编码规则

GB/T 13000 信息技术 通用多八位编码字符集（UCS）

【条文解读】

本标准所引用的 3 项国家标准，对本标准的实现具有重要影响。

（1）GB/T 2312—1980《信息交换用汉字编码字符集 基本集》

该标准规定了 6763 个汉字以及一般符号、序号、数字、拉丁字母、日文假名、希腊字母、俄文字母、汉语拼音符号、汉语注音字母等图形字符，共计 7445 个字符。该标准的字汇范围较为有限，仅收录了常用的一些汉字，主要满足不具备交换功能的设备对于中文信息处理的需求。本标准双字节部分收录了该标准内全部图形字符。

（2）GB/T 11383—1989《信息处理 信息交换用八位代码结构和编码规则》

该标准的字汇包括 52 个基本的小写和大写拉丁字母，以及音调字母、专用拉丁字母和几种非拉丁字母。本标准单字节部分收录了该标准的 97 个字符，并且每个字符的码位均与该标准保持一致。

（3）GB/T 13000《信息技术 通用多八位编码字符集（UCS）》

GB/T 13000 等同采用国际标准 ISO/IEC 10646。该标准对全世界古今文字和图形符号统一编码，解决多文种统一处理的问题。该标准收录了我国使用的现代汉字和蒙古文、藏文等多种现行少数民族文字，也收录了八思巴文、西夏文等多种我国古文字。本标准双字节和四字节部分收录了该标准的一些字符，并且在本标准代码表中给出了每一个字符对应的 GB/T 13000 码位。

本标准发布时，GB/T 13000 的有效版本是 2010 年版，对应于国际标准 ISO/IEC 10646: 2003，而本标准引用了大量来自国际标准 ISO/IEC 10646: 2017 的字符，如 7473 个汉字（CJK 统一汉字扩充 F 部分）和 13 个蒙古文 BIRGA 符号（蒙古文补充部分）。虽然 GB/T 13000 的修订已经列入标准编制人员的工作

计划，但由于工作量和工作程序等原因，无法在本标准发布前完成。因此，为避免误解，本标准在引用 GB/T 13000 时未注日期。

2.3 术语和定义的解读

【标准条文】

3 术语和定义

下列术语和定义适用于本文件。

3.1

字符 character

供组织、控制或表示数据用的元素集合中的一个元素。

3.2

编码字符 coded character

字符及其编码表示。

3.3

用户自定义区 private use area

使用符合本标准的产品的使用者可以自行规定的区域。

3.4

字汇 repertoire

用编码字符集表示的一个指定的字符集合。

3.5

保留区 reserved zone

留作未来国家标准规定的区域。

【条文解读】

——字符

在不同的语境中，“字符”一词具有不同的含义。在书写系统中，字符就是用于传递信息的具体图形符号。这些图形符号可以是内含语音的文字，也可以是无语音的图形或抽象符号（如交通标志）。而在 GB 18030 中，字符可以是用于自然语言书写形式的基本单位（如阿拉伯数字、汉字或拉丁字母、标点符号、空格等），也可以是用于信息处理系统的非图形单位（如控制字符）。

——编码字符

即经过编码的字符。GB 18030 规定的字符都是编码字符。

——用户自定义区

GB 18030 定义的用户是“最终用户”，即产品的最终使用者。例如，操作系统开发者不是最终用户，基于操作系统的应用软件开发者也不是最终用户。因此，任何投放市场的公开而面向无差别对象销售的产品，都不应占用标准规定的用户自定义区。反之，不公开销售而面向特定对象销售的产品，即用户定制产品，不受此限制。需要指出的是，正在市场上销售的某些“XX 通信集团有限公司定制移动通信终端”不属于上述用户定制产品，因为这些产品“公开而面向无差别对象销售”。

非最终用户而错误占用了用户自定义区的例子如：为迎合大量用户对表情符号（日文：絵文字/えもじ emoji，如下图所示）的喜爱，某些产品（如移动电话）将这些符号置于用户自定义区，不符合本标准的规定。在此提出一种解决办法，供产品开发者参考：未经编码的字符（如上述表情符号）可以升级安装包形式供最终用户选择下载安装，但应具备卸载功能，并应在下载或安装前向最终用户展示免责声明。

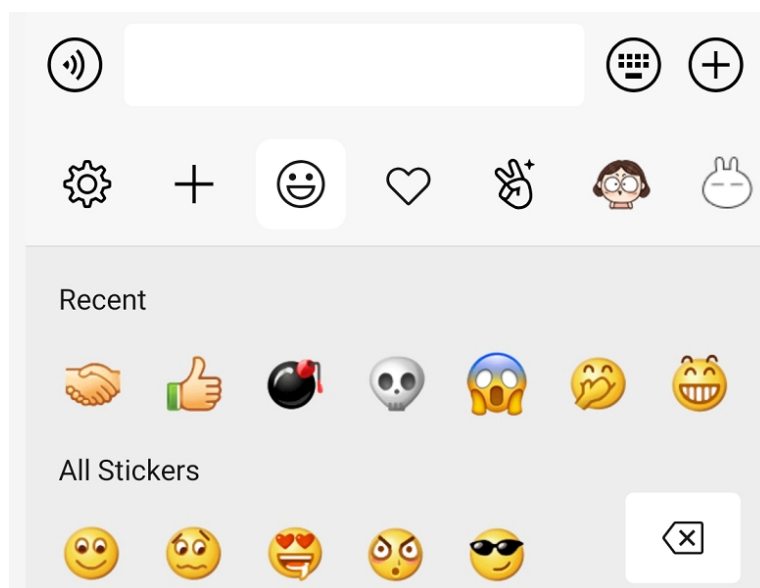


图 1 某品牌移动通信终端使用的表情符号²

——字汇

即字符的集合。例如，GB 18030 的汉字字汇共收录了来自国际标准 ISO/IEC

² 本图仅用于示意表情符号，不表示该产品自带的表情符号占用了用户自定义区。



10646 (idt GB/T 13000) 的 CJK 统一汉字 (及其扩充) 共 87875 个。另外还有 12 个 CJK 兼容汉字。

关于 CJK 统一汉字、CJK 兼容汉字的说明, 详见第二章第七节“关于码位分配的补充说明”及第五章第四节“CJK 统一汉字及其扩充的介绍”。

——保留区

GB 18030 未来修订时可能重新规定的区域, 任何使用当前版本的产品开发者均不得占用。

2.4 标准收录的字汇

【标准条文】

4 字汇

4.1 概述

本标准收录的字符以单字节、双字节或四字节编码。

4.2 单字节部分

单字节的部分收录了 GB/T 11383—1989 的 0x00 到 0x7F 全部 128 个字符。

4.3 双字节部分

双字节的部分收录了 GB/T 2312—1980 中的全部图形字符、GB/T 13000 中的 CJK 统一汉字以及部分图形字符。双字节部分的字符按照附录 A 的规定。其中, 表意文字描述符的图形、代码位置和功能描述见附录 B。

注: GB/T 13000 对中国、日本、韩国、越南等国家和地区使用的汉字统一编码。具备独有抽象字形的汉字, 赋予单独的代码位置; 来源不同而抽象字形相同的汉字, 则赋予共用的代码位置。经过编码的汉字称为 CJK 统一汉字 (CJK Unified Ideographs), 其中 CJK 表示中国、日本、韩国。

4.4 四字节部分

四字节的的部分收录了上述双字节字符之外的 GB/T 13000 的 66 个 CJK 统一汉字 (9FA6 到 9FEF, 不包括 9FB4 到 9FBB 的 8 个字符)、CJK 统一汉字扩充 A、CJK 统一汉字扩充 B、CJK 统一汉字扩充 C、CJK 统一汉字扩充 D、CJK 统一汉字扩充 E、CJK 统一汉字扩充 F 和已经在 GB/T 13000 中编码的少数民族文字的字符。四字节部分的字符按照附录 C 的规定。

【条文解读】

2.4.1 概述

第 4 章分别规定了标准单字节编码部分、双字节编码部分、四字节编码部分

的字汇。单字节编码部分的字汇来自于 GB/T 11383—1989，双字节编码部分和四字节编码部分的字汇都来自于 GB/T 13000。GB 18030 规定的字符和 GB/T 13000 规定的字符一一对应，为采用这两种不同编码体系的信息处理系统之间的兼容和信息转换提供基础。

2.4.2 单字节部分的字汇

单字节部分的字汇与 GB/T 11383—1989 的字汇完全一致，其编码与 GB/T 11383—1989 的编码完全一致，详见 GB 18030 的第 7 章。单字节部分的总字符数为 97 个。

2.4.3 双字节部分的字汇

双字节部分包括常用的数字、字母、日文假名、标点、符号、汉语拼音、注音符、CJK 统一汉字、CJK 兼容汉字和一些汉字构件。双字节部分的总字符数为 21882 个。

2.4.4 四字节部分的字汇

四字节部分的字汇包括 CJK 统一汉字和 10 种我国少数民族现行文字的字符，全部来自于国际标准 ISO/IEC 10646: 2017。四字节部分的总字符数为 72881 个。

上文所说的 10 种少数民族现行文字的字符，包括蒙古文（包括传统蒙古文、满文、托忒文、锡伯文、阿礼嘎礼字）、藏文、维哈柯文（包括维吾尔文、哈萨克文、柯尔克孜文）、朝鲜文、规范彝文、德宏傣文、西双版纳新傣文、西双版纳老傣文、傈僳文、滇东北苗文。其中，维吾尔、哈萨克、柯尔克孜文均使用阿拉伯字母，统一编码；传统蒙古文、满文、托忒文、锡伯文、阿礼嘎礼字均使用蒙古文字母，统一编码。以上统计为 10 种。

由于字形特点和使用方式，方块壮字纳入 CJK 统一汉字，分布于本标准的 0x9839B632—0x9933FE33（CJK 统一汉字扩充 E）和 0x99348138—0x9939F730（CJK 统一汉字扩充 F）区间。

标准第 7 章“码位分配”详细规定了各文种的名称和编码位置。

2.5 标准总体编码结构

【标准条文】



5 总体结构

正文中凡数字前标有 0x 的表示采用十六进制, 未标有 0x 的表示采用十进制。
附录中凡编码的表示均采用十六进制, 其他数字的表示均采用十进制。

单字节部分采用 GB/T 11383—1989 的编码结构, 使用 0x00 至 0x7F 码位。

双字节部分采用两个八位二进制位串表示一个字符, 其首字节码位从 0x81 至 0xFE, 尾字节码位分别是 0x40 至 0x7E 和 0x80 至 0xFE。

四字节部分采用 GB/T 11383—1989 未采用的 0x30 到 0x39 作为对双字节编码扩充的后缀, 编码范围为 0x81308130 到 0xFE39FE39。四字节字符的第一个字节编码范围为 0x81 至 0xFE; 第二个字节编码范围为 0x30 至 0x39; 第三个字节编码范围为 0x81 至 0xFE; 第四个字节编码范围为 0x30 至 0x39。即,

0x81308130 至 0x81308139;

0x81308230 至 0x81308239;

.....

0x8130FE30 至 0x8130FE39;

0x81318130 至 0x81318139;

.....

0x8131FE30 至 0x8131FE39;

.....

0x82308130 至 0x82308139;

.....

0x8230FE30 至 0x8230FE39;

.....

0xFE308130 至 0xFE308139;

.....

0xFE39FE30 至 0xFE39FE39。

总体结构图 (略)

【条文解读】

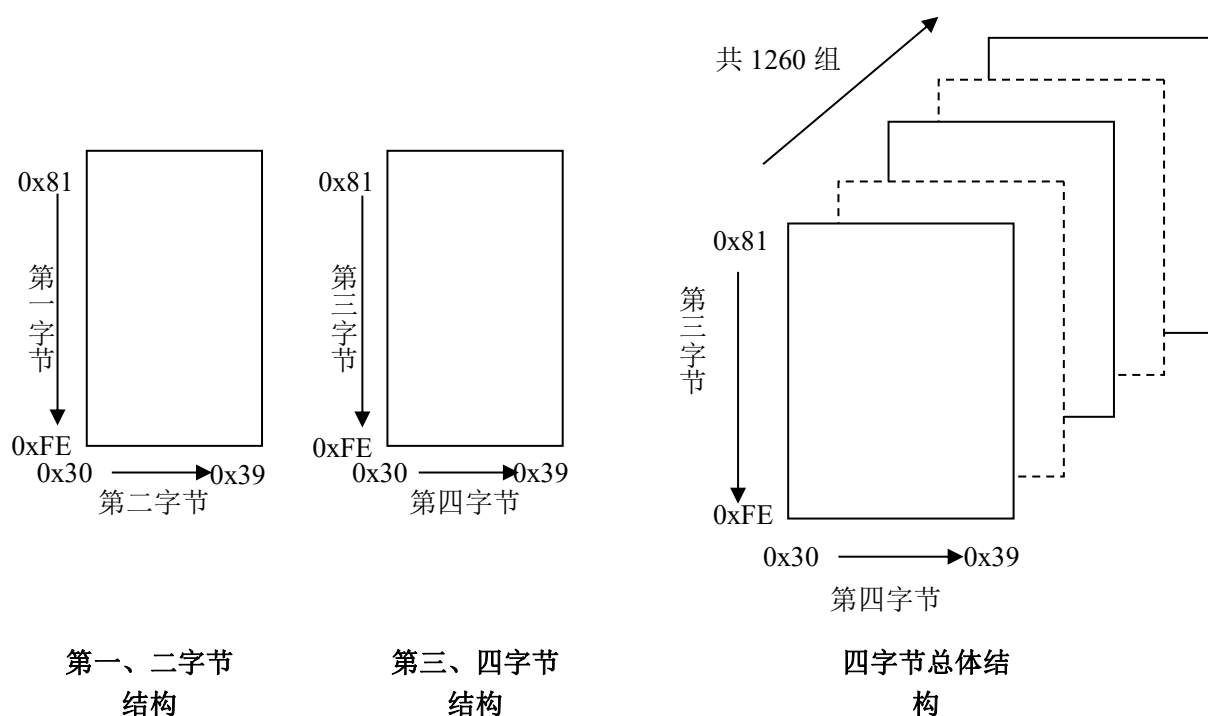
本标准规定的编码空间是在已废止的指导性技术文件《汉字扩展内码规范 (GBK) 1.0 版》基础上的扩充。本标准采用单字节、双字节和四字节三种方式对字符编码。

单字节部分采用 GB 11383—1989 《信息处理 信息交换用八位代码结构和编码规则》（idt ISO 4873: 1986）的编码结构与规则，使用 0×00 至 0×7F 码位。

双字节部分，首字节码位从 0×81 至 0×FE，尾字节码位分别是 0×40 至 0×7E 和 0×80 至 0×FE。这部分完全采用了 GBK 的编码结构，但用户自定义区码位与 GB/T 13000 码位的对应关系有所调整。

四字节部分采用 GB/T 11383 未采用的 0×30 到 0×39 作为对双字节编码扩充的后缀，这样扩充的四字节编码，其范围为 0×81308130 到 0×FE39FE39。其中，第一、三个字节编码码位均为 0×81 至 0×FE，第二、四个字节编码码位均为 0×30 至 0×39。以第一和第二字节确定平面顺序，从 0×8130 平面至 0×FE39 平面，共 1260 个平面；以第三字节和第四字节确定每个平面上的码位，从 0×8130 码位至 0×FE39 码位，每个平面有 1260 个码位；1260 个平面上共有 1587600 个码位。

本标准四字节结构的设计参考了 GB/T 13000—2010 中的平面、码位概念。这里摘录两个标准的相关图示和文字，以便读者比较。

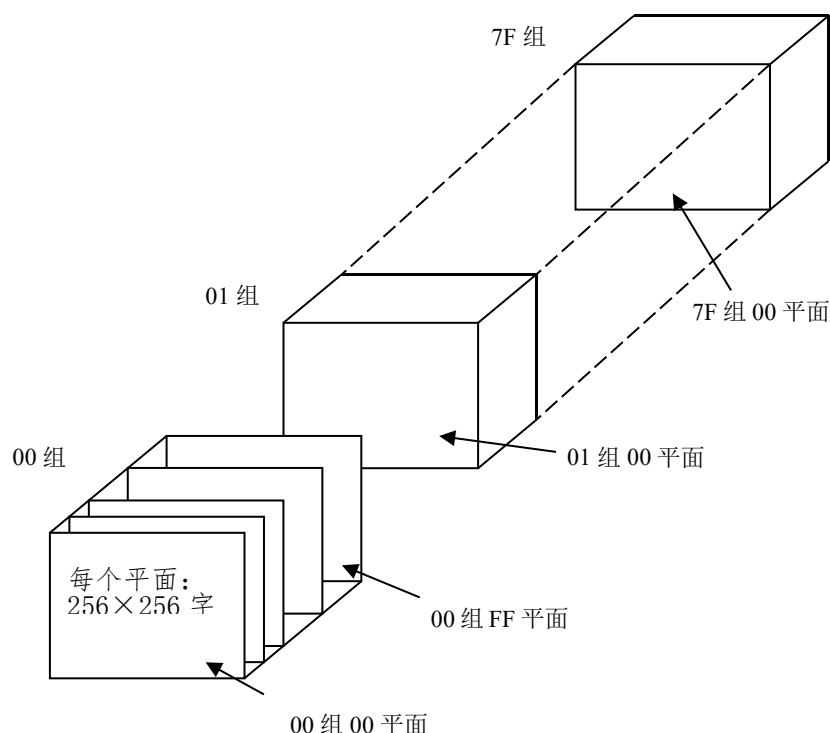


GB 18030—2022 图 1 总体结构图

下面是 GB/T 13000-2010 的 6.1 《结构》的相关部分：



整个编码字符集应被表达为包含 128 个组，其中每组有 256 个平面。每一平面应被视为含有 256 行字符，每行 256 个字位。在表示平面内容（如图 2）的代码表中，水平轴应表示最低八位，越向左其值越小；而纵轴应表示较高八位，越向上其值越小。



GB/T 13000-2010 图 1 通用多八位编码字符集的全部编码空间

2.6 字符的排列顺序

【标准条文】

6 字符的排列顺序

6.1 单字节部分字符的排列顺序

单字节部分字符按照 GB/T 11383—1989 中相应字符的顺序排列。

6.2 双字节部分字符的排列顺序

双字节部分字符排列顺序应符合附录 A。

6.3 四字节部分字符的排列顺序

自 0x81308130 至 0x8439FE39 共 50400 个码位，对应双字节部分未包括的所有 GB/T 13000 基本多文种平面的字符，按照 GB/T 13000 基本多文种平面相应字符的顺序排列。

自 0x90308130 至 0xE339FE39 共 1058400 个码位用于对应 GB/T 13000 的 16 个辅助平面, 字符排列顺序完全遵照 GB/T 13000 的 16 个辅助平面的相应码位顺序依次排列。

四字节部分字符表应符合附录 C。

【条文解读】

单字节编码的字符包括各种标点、符号、阿拉伯数字和拉丁字母以及控制字符。其中, 阿拉伯数字按照增序排列; 拉丁字母的大小写形式分别按照字母表顺序排列。

双字节 1 区编码的阿拉伯数字按照增序排列; 拉丁字母、希腊字母、斯拉夫字母的大小写形式分别按照相应文种的字母表顺序排列; 日文平假名和片假名分别按照五十音顺序排列。

双字节 2 区编码的汉字按照 GB/T 2312 所收汉字的顺序排列。

双字节 3 区编码的汉字和双字节 4 区编码的汉字按照 GB/T 13000 中相应汉字的顺序收入本标准。

四字节编码的字符全部来自国际标准 ISO/IEC 10646 (idt GB/T 13000), 按照 GB/T 13000 码位和本标准码位对应关系, 收入本标准。

2.7 关于码位分配的补充说明

【标准条文】

7 码位分配

标准条文 (略)。

【条文解读】

标准条文第 7 章《码位分配》的表述比较清楚, 请读者参阅原文。在此仅做一些补充说明, 供读者参考。

2.7.1 CJK 统一汉字

本标准中, 数量最大的字汇是 CJK 统一汉字 (CJK UNIFIED IDEOGRAPHS)。关于 CJK 统一汉字的更多信息, 参见本文件第五章第四节“CJK 统一汉字及其扩充的介绍”。

本标准共收录 87875 个 CJK 统一汉字, 与 GB/T 13000 的相应字块关系如下:



- CJK 统一汉字：20976 个
- CJK 统一汉字扩充 A：6582 个
- CJK 统一汉字扩充 B：42711 个
- CJK 统一汉字扩充 C：4149 个
- CJK 统一汉字扩充 D：222 个
- CJK 统一汉字扩充 E：5762 个
- CJK 统一汉字扩充 F：7473 个

2.7.2 CJK兼容汉字

由于国家、民族、文化、历史等原因，某些国家标准/地区标准对同一个汉字进行了重复编码。CJK 兼容汉字主要是为了与这些现存编码字符集进行双向代码转换，而不会丢失信息。如非为了满足本标准和这些国家标准/地区标准的双向转换的需要，不得使用兼容汉字。

CJK 兼容汉字在本标准置于双字节 4 区，其字形、代码位置及其与 GB/T 13000 的对应关系如下表白色部分所示：

FE	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	𪛗 FA0C	𪛘 FA0D	𪛚 FA0E	𪛜 FA0F	𪛞 FA11	𪛠 FA13	𪛡 FA14	𪛣 FA18	𪛥 FA1F	𪛧 FA20	𪛩 FA21	𪛫 FA23	𪛭 FA24	𪛯 FA27	𪛱 FA28	𪛳 FA29
5	𪛵 2E81	𪛷 E816	𪛹 E817	𪛻 E818	𪛽 2E84	𪛿 3473	𪛻 3447	𪛽 2E88	𪛿 2E8B	𪛻 9FB4	𪛽 359E	𪛿 361A	𪛻 360E	𪛽 2E8C	𪛿 2E97	𪛻 396E
6	𪛵 3918	𪛷 9FB5	𪛹 39CF	𪛻 39DF	𪛽 3A73	𪛿 39D0	𪛻 9FB6	𪛽 9FB7	𪛿 3B4E	𪛻 3C6E	𪛽 3CE0	𪛿 2EA7	𪛻 E831	𪛽 9FB8	𪛿 2EAA	𪛻 4056
7	𪛵 415F	𪛷 2EAE	𪛹 4337	𪛻 2EB3	𪛽 2EB6	𪛿 2EB7	𪛻 E83B	𪛽 43B1	𪛿 43AC	𪛻 2EBB	𪛽 43DD	𪛿 44D6	𪛻 4661	𪛽 464C	𪛿 9FB9	
8	𪛵 4723	𪛷 4729	𪛹 477C	𪛻 478D	𪛽 2ECA	𪛿 4947	𪛻 497A	𪛽 497D	𪛿 4982	𪛻 4983	𪛽 4985	𪛿 4986	𪛻 499F	𪛽 499B	𪛿 49B7	𪛻 49B6
9	𪛵 9FBA	𪛷 E855	𪛹 4CA3	𪛻 4C9F	𪛽 4CA0	𪛿 4CA1	𪛻 4C77	𪛽 4CA2	𪛿 4D13	𪛻 4D14	𪛽 4D15	𪛿 4D16	𪛻 4D17	𪛽 4D18	𪛿 4D19	𪛻 4DAE
A	𪛵 9FBB															

需要说明的是，GB 18030 第三版删除了 2000 年版和 2005 年版收录的 9 个 CJK 兼容汉字。这 9 个 CJK 兼容汉字分别具有明确对应的 CJK 统一汉字，对应关系如下：



GB 18030 第三版的码位	对应 GB/T 13000 兼容汉字的码位	对应 GB/T 13000 兼容汉字的字形	对应 GB/T 13000CJK 统一汉字的码位	对应 GB/T 13000CJK 统一汉字的字形
FD9C	F92C	郎	90CE	郎
FD9D	F979	凉	51C9	凉
FD9E	F995	季	5B63	季
FD9F	F9E7	裏	88CF	裏
FDA0	F9F1	隣	96A3	隣
FE40	FA0C	兀	5140	兀
FE41	FA0D	殼	55C0	殼
FE47	FA18	礼	793C	礼
FE49	FA20	蘊	8612	蘊

2.7.3 汉字部首

本标准中的汉字部首包括康熙部首和补充 CJK 部首，合计为 228 个。

《康熙字典》³规定的 214 个部首，称为康熙部首（KANGXI RADICALS）。康熙部首在本标准置于四字节编码区，其字形、代码位置及其与 GB/T 13000 的对应关系如下表所示：

8139

	30	31	32	33	34	35	36	37	38	39
8B	2EFE	2EFF	一 2F00	丨 2F01	丶 2F02	丿 2F03	乙 2F04	丿 2F05	二 2F06	亅 2F07
8C	人 2F08	儿 2F09	入 2F0A	八 2F0B	冂 2F0C	冂 2F0D	冂 2F0E	几 2F0F	凵 2F10	刀 2F11
8D	力 2F12	勹 2F13	匕 2F14	匚 2F15	匚 2F16	十 2F17	卜 2F18	冂 2F19	厂 2F1A	厶 2F1B
8E	又 2F1C	口 2F1D	口 2F1E	土 2F1F	士 2F20	夂 2F21	夂 2F22	夕 2F23	大 2F24	女 2F25
8F	子 2F26	乚 2F27	寸 2F28	小 2F29	尢 2F2A	尸 2F2B	屮 2F2C	山 2F2D	巛 2F2E	工 2F2F
90	己 2F30	巾 2F31	干 2F32	幺 2F33	广 2F34	爻 2F35	升 2F36	弋 2F37	弓 2F38	厶 2F39
91	彡 2F3A	彳 2F3B	心 2F3C	戈 2F3D	戶 2F3E	手 2F3F	支 2F40	支 2F41	文 2F42	斗 2F43
92	斤 2F44	方 2F45	无 2F46	日 2F47	曰 2F48	月 2F49	木 2F4A	欠 2F4B	止 2F4C	歹 2F4D
93	殳 2F4E	母 2F4F	比 2F50	毛 2F51	氏 2F52	气 2F53	水 2F54	火 2F55	爪 2F56	父 2F57

³ 《康熙字典》选用北京中华书局影印本（1958 年 1 月第 1 版）。



94	爻 2F58	爿 2F59	片 2F5A	牙 2F5B	牛 2F5C	犬 2F5D	玄 2F5E	玉 2F5F	瓜 2F60	瓦 2F61
95	甘 2F62	生 2F63	用 2F64	田 2F65	疋 2F66	疒 2F67	死 2F68	白 2F69	皮 2F6A	皿 2F6B
96	目 2F6C	矛 2F6D	矢 2F6E	石 2F6F	示 2F70	肉 2F71	禾 2F72	穴 2F73	立 2F74	竹 2F75
97	米 2F76	糸 2F77	缶 2F78	网 2F79	羊 2F7A	羽 2F7B	老 2F7C	而 2F7D	耒 2F7E	耳 2F7F
98	聿 2F80	肉 2F81	臣 2F82	自 2F83	至 2F84	臼 2F85	舌 2F86	舛 2F87	舟 2F88	艮 2F89
99	色 2F8A	艸 2F8B	虍 2F8C	虫 2F8D	血 2F8E	行 2F8F	衣 2F90	兩 2F91	見 2F92	角 2F93
9A	言 2F94	谷 2F95	豆 2F96	豕 2F97	豸 2F98	貝 2F99	赤 2F9A	走 2F9B	足 2F9C	身 2F9D
9B	車 2F9E	辛 2F9F	辰 2FA0	辵 2FA1	邑 2FA2	酉 2FA3	采 2FA4	里 2FA5	金 2FA6	長 2FA7
9C	門 2FA8	阜 2FA9	隶 2FAA	隹 2FAB	雨 2FAC	青 2FAD	非 2FAE	面 2FAF	革 2FB0	韋 2FB1
9D	韭 2FB2	音 2FB3	頁 2FB4	風 2FB5	飛 2FB6	食 2FB7	首 2FB8	香 2FB9	馬 2FBA	骨 2FBB
9E	高 2FBC	髟 2FBD	鬥 2FBE	鬯 2FBF	鬲 2FC0	鬼 2FC1	魚 2FC2	鳥 2FC3	鹵 2FC4	鹿 2FC5
9F	麥 2FC6	麻 2FC7	黃 2FC8	黍 2FC9	黑 2FCA	黽 2FCB	黽 2FCC	鼎 2FCD	鼓 2FCE	鼠 2FCF
A0	鼻 2FD0	齊 2FD1	齒 2FD2	龍 2FD3	龜 2FD4	龠 2FD5				

为了满足各方对康熙部首之外汉字部首的使用需要，GB/T 13000 又增收了 116 个汉字部首，称为补充 CJK 部首（CJK RADICALS SUPPLEMENT）。由于绝大部分补充 CJK 部首与已编码的汉字在字形上完全一致，为避免误用造成混乱，本标准仅收录了 14 个曾在 GBK 中编码的补充 CJK 部首，并未收录其余的补充 CJK 部首。GB 18030 收录的 14 个补充 CJK 部首置于双字节 4 区，其字形、代码位置及其与 GB/T 13000 的对应关系如下表白色部分所示：



FE	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4			夔	垓	崎	梔	榉		藹		虍	尅	返	銑	鋅	隄
	FA0C	FA0D	FA0E	FA0F	FA11	FA13	FA14	FA18	FA1F	FA20	FA21	FA23	FA24	FA27	FA28	FA29
5	𠂇				𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇
	2E81	E816	E817	E818	2E84	3473	3447	2E88	2E8B	9FB4	359E	361A	360E	2E8C	2E97	396E
6	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇
	3918	9FB5	39CF	39DF	3A73	39D0	9FB6	9FB7	3B4E	3C6E	3CE0	2EA7	E831	9FB8	2EAA	4056
7	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇
	415F	2EAE	4337	2EB3	2EB6	2EB7	E83B	43B1	43AC	2EBB	43DD	44D6	4661	464C	9FB9	
8	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇
	4723	4729	477C	478D	2ECA	4947	497A	497D	4982	4983	4985	4986	499F	499B	49B7	49B6
9	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇
	9FBA	E855	4CA3	4C9F	4CA0	4CA1	4C77	4CA2	4D13	4D14	4D15	4D16	4D17	4D18	4D19	4DAE
A	𠂇															
	9FBB															

2.7.4 少数民族文字

本标准规定的我国少数民族文字的各个文种，一般整体收录 GB/T 13000 相关字块（BLOCK）中的字符，但对朝鲜文和维吾尔文、哈萨克文、柯尔克孜文（以下简称维哈柯文）依据相关现行国家标准进行了选择性收录。这些少数民族文字字符与国家标准的关系见下表：

文种	字符数	字汇来源
朝鲜文	3551	GB/T 12052-1989《信息交换用朝鲜文字编码字符集》
维哈柯文	185	GB/T 21669-2008《信息技术 维吾尔文、哈萨克文、柯尔克孜文编码字符集》

本标准仅收录了 GB/T 13000 中可以与 GB/T 12052-1989《信息交换用朝鲜文字编码字符集》对应的 3551 个朝鲜文字符。包括 69 个朝鲜文字母、51 个朝鲜文兼容字母和 3431 个朝鲜文音节。

本标准共收录 185 个维哈柯文字符，均来自 GB/T 21669-2008《信息技术 维吾尔文、哈萨克文、柯尔克孜文编码字符集》。维哈柯文的字符收录于 GB/T 13000 中的阿拉伯文相关字块，但未曾逐个指明所属文种。GB/T 21669-2008 从 GB/T 13000 中的阿拉伯文相关字块选取维哈柯文所用字符，分别说明。185 个字符中的部分字符，为维吾尔文、哈萨克文、柯尔克孜文中的两种或三种所共用。

2.8 部分字符和代码的说明

【标准条文】

8 部分字符和代码的说明

相对于 GB 18030-2005，部分代码位置上的字形和/或所对应的 GB/T 13000 代码位置在本文件中进行了调整（见附录 D）。

【条文解读】

参阅本标准附录 D（资料性）《部分字符和代码的说明》的说明。附录 D 用表格形式列出了调整过的 18 个 GB/T 13000 代码位置、在双字节区删除的 6 个汉字构件和 9 个 CJK 兼容汉字。

2.9 三个实现级别的界定

【标准条文】

9 实现的级别

9.1 通则

本文件规定三个实现级别。符合相应实现级别的系统软件产品，应提供相应实现级别范围内全部字符的输入输出功能。

9.2 实现级别 1

实现级别 1 支持本文件的单字节编码部分、双字节编码部分和四字节编码部分的 CJK 统一汉字（即 0x82358F33—0x82359636）和 CJK 统一汉字扩充 A（即 0x8139EE39—0x82358738）。

任何本文件适用的产品均应满足实现级别 1 的要求。

注：根据软件应用需要，实现级别 1 还可选择支持表 3 列出的任何一种或多种非汉字文种。

9.3 实现级别 2

实现级别 2 包含实现级别 1。此外，实现级别 2 还支持《通用规范汉字表》中的没有包含在实现级别 1 之内的编码汉字。《通用规范汉字表》所收汉字在本文件中的代码位置和字形，见附录 E。

系统软件及支撑软件，应满足实现级别 2 的要求。

注：系统软件及支撑软件包括但不限于操作系统、数据库管理系统、中间件（软件产品分类的信息见 GB/T 36475）。

9.4 实现级别 3

实现级别 3 包含实现级别 2。此外，实现级别 3 还支持本文件规定的全部汉字及表 3 中的康熙部首。

用于政务服务和公共服务的产品应满足实现级别 3 的要求。

注：政务服务和公共服务行业包括但不限于铁路运输业、道路运输业、水上运输业、航空运输业、多式联运和运输代理业、邮政业、货币金融服务、保险业、土地管理业、卫生、国家机构、社会保障等（行业分类的信息见 GB/T 4754）。

【条文解读】



2.9.1 实现级别1

实现级别 1 是对具备中文信息处理功能的产品的基本要求，共包含 28575 个字符，其中汉字和汉字部首数为 27584 个。字汇基本等同于本标准上一版，即 GB 18030—2005 前言部分的强制性要求⁴，但在此基础上增加了 66 字。符合本标准上一版强制性要求的产品在升级字库后即可达到本标准实现级别 1 的要求。

实现级别 1 的规定，既考虑了符合 GB 18030—2005 的已上市产品对符合新标准兼容性需要，也考虑了多数产品采用 GB/T 13000 内核而因编码框架限制带来字汇大小受限的现实问题。后者是指 GB/T 13000 的基本多文种平面（BMP）字符可以双字节处理，而辅助多文种平面（SMP）字符则须以四字节或其它较为复杂的方式处理，技术难度提高。本标准实现级别 1 规定的汉字与 GB/T 13000 的 BMP 中的汉字一一对应。换言之，凡采用 GB/T 13000 内核的产品，只要支持 BMP，即支持本标准实现级别 1。

2.9.2 实现级别2

实现级别 2 在字汇上相较实现级别 1 增加了 196 个汉字，总字符数为 28771 个，其中汉字和汉字部首数为 27780 个。该级别的要义在于满足实现《通用规范汉字表》的要求。《通用规范汉字表》由教育部组织编制，共收录 8105 个汉字，该表于 2013 年由国务院批准发布。⁵

对于采用 GB/T 13000 内核的产品，完整支持《通用规范汉字表》的汉字，在技术上有较高要求，因为该字表规定的 196 个汉字（3 个二级汉字和 193 个三级汉字）位于 GB/T 13000 的 SMP 中。

2.9.3 实现级别3

实现级别 3 包括标准内定义的全部 88115 个汉字和汉字部首，以及单字节编码部分和双字节 1 区、5 区的各类图形符号，总计 89106 个字符。实现级别 3 的要求，主要面向行业应用，特别是 GB/T 4754《国民经济行业分类》所规定的与国计民生密切相关的行业，如：

S 公共管理、社会保障和社会组织

92 国家机构

94 社会保障

⁴ GB 18030—2005 的前言声明：本标准的单字节编码部分、双字节编码部分和四字节编码部分的 CJK 统一汉字扩充 A（即 0x8139EE39—0x82358738）部分为强制性标准。

⁵ 《国务院关于公布《通用规范汉字表》的通知》（国发〔2013〕23 号）

注：S 为门类编号，92、93 为大类编号。

门类的划分，详见 GB/T 4754。

第三章 标准五个附录的解读

3.1 附录 A（规范性）双字节字符表

【标准条文】

A.1 表的内容

本表给出双字节编码字符的字形及代码，同时给出了该字符对应的 GB/T 13000 代码位置。

A.2 表的说明

示例如下：

（略）

【条文解读】

本标准中的双字节编码字符的代码与字形均在双字节字符表中列出。双字节字符表按照字符代码顺序排列。表左上角单元格显示的 2 位数字是该表内所有字符代码相同的第一字节，最左侧一列是代码的第二字节高位，最上方一行是代码的第二字节低位。当读取位于双字节区的某字符的 GB 18030 代码时，其顺序为首先第一字节、其次第二字节高位、最后第二字节低位。两个字节依次连接，组成了该字符的 GB 18030 代码。

对于本标准收录的汉字和部分少数民族文字，在其代码对应的单元格内显示该字符的字形，并在字形下方给出该字符对应的 GB/T 13000 代码。其余字符因不在本标准文种收录范围，不显示字形，而仅在单元格中给出该字符对应的 GB/T 13000 代码。

GB 18030-2022 中所示字形仅作为该图形字符的参考字形。汉字及其他图形字符的具体字形应符合相关字型国家标准的规定。

下面举例列出本标准图 A.1 中部分字符的字形、代码及其对应的 GB/T 13000 代码，以利读者理解。



81	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	𠂇 4E02	𠂈 4E04	𠂉 4E05	𠂊 4E06	𠂋 4E0F	𠂌 4E12	𠂍 4E17	𠂎 4E1F	𠂏 4E20	𠂐 4E21	𠂑 4E23	𠂒 4E26	𠂓 4E29	𠂔 4E2E	𠂕 4E2F	𠂖 4E31
5	𠂗 4E33	𠂘 4E35	𠂙 4E37	𠂚 4E3C	𠂛 4E40	𠂜 4E41	𠂝 4E42	𠂞 4E44	𠂟 4E46	𠂠 4E4A	𠂡 4E51	𠂢 4E55	𠂣 4E57	𠂤 4E5A	𠂥 4E5B	𠂦 4E62

上图是本标准图 A.1 的节选，其中各字符的字形、代码及其对应的 GB/T 13000 代码分别如下表所示：

字形	GB 18030 代码	GB/T 13000 代码	字形	GB 18030 代码	GB/T 13000 代码
𠂇	8140	4E02	𠂗	8150	4E33
𠂈	8141	4E04	𠂘	8151	4E35
𠂉	8142	4E05	𠂙	8152	4E37
𠂊	8143	4E06	𠂚	8153	4E3C
𠂋	8144	4E0F	𠂛	8154	4E40
𠂌	8145	4E12	𠂜	8155	4E41
𠂍	8146	4E17	𠂝	8156	4E42
𠂎	8147	4E1F	𠂞	8157	4E44
𠂏	8148	4E20	𠂟	8158	4E46
𠂐	8149	4E21	𠂠	8159	4E4A
𠂑	814A	4E23	𠂡	815A	4E51
𠂒	814B	4E26	𠂢	815B	4E55
𠂓	814C	4E29	𠂣	815C	4E57
𠂔	814D	4E2E	𠂤	815D	4E5A
𠂕	814E	4E2F	𠂥	815E	4E5B
𠂖	814F	4E31	𠂦	815F	4E62

3.2 附录 B（规范性）表意文字描述符

【标准条文】

本附录收录了 GB/T 13000 规定的表意文字描述符 13 个，并给出了它们的图形、代码位置和功能描述。具体内容见表 B.1。

表 B.1 表意文字描述符的图形、代码位置和功能描述（略）

【条文解读】



下面引用 GB/T 13000—2010 附录 F 中（资料性附录）《替代格式字符》中有关表意文字描述符的文字，并对其组合方式举例略作说明。

表意文字描述序列（IDS）用于描述尚未在标准中规定的表意文字字符。IDS 由表意文字描述符及其随后的固定数目的描述部件（DC）组成。DC 可以是下列之一：

- 编码汉字；
- 编码部首；
- 另一个 IDS。

表 F.1 表意文字描述符的属性

字符名称： IDEOGRAPHIC DESCRIPTION CHARACTER ...	DC 数目	IDS缩略语与句 法	DC的相关 位置	IDS示例	IDS示例 显现
LEFT TO RIGHT	2	IDC-LTR D ₁ D ₂		𠂇 母	𠂇
ABOVE TO BELOW	2	IDC-ATB D ₁ D ₂		𠂇 天	𠂇
LEFT TO MIDDLE AND RIGHT	3	IDC-LMR D ₁ D ₂ D ₃		讠 言 子	讠
ABOVE TO MIDDLE AND BELOW	3	IDC-AMB D ₁ D ₂ D ₃		目 从 从 日	目
FULL SURROUND	2	IDC-FSD D ₁ D ₂		回 口 巷	回
SURROUND FROM ABOVE	2	IDC-SAV D ₁ D ₂		門 卜	門
SURROUND FROM BELOW	2	IDC-SBL D ₁ D ₂		山 士	山
SURROUND FROM LEFT	2	IDC-SLT D ₁ D ₂		匚 虎	匚



SURROUND FROM UPPER LEFT	2	IDC-SUL D ₁ D ₂			靡
SURROUND FROM UPPER RIGHT	2	IDC-SUR D ₁ D ₂			匄
SURROUND FROM LOWER LEFT	2	IDC-SLL D ₁ D ₂			这
OVERLAID	2	IDC-OVL D ₁ D ₂			巫

注：D₁和D₂互相嵌入。本图表不表示D₁在左上角或D₂在右下角。

3.3 附录 C（规范性）四字节字符表

【标准条文】

C.1 表的内容

本表部分给出了汉字和部分少数民族文字的字形及代码，同时还给出了该字符对应的 GB/T 13000 代码位置，其它部分的字形从略。

C.2 表的说明

示例如下：

（略）

【条文解读】

本标准中的四字节编码字符的代码与字形均在四字节字符表中列出。四字节字符表按照字符代码顺序排列。表上方居中位置显示的 4 位数字是该表内所有字符代码相同的第一和第二字节，最左侧一列是代码的第三字节，最上方一行是代码的第四字节。当读取位于四字节区的某字符在本标准中的代码时，其顺序为首先第一和第二字节、其次第三字节、最后第四字节。四个字节依次连接，组成了该字符的 GB 18030 代码。

对于本标准收录的汉字和部分少数民族文字，在其代码对应的单元格内显示该字符的字形，并在字形下方给出该字符对应的 GB/T 13000 代码。其余字符因不在本标准文种收录范围，不显示字形，而仅在单元格中给出该字符对应的 GB/T 13000 代码。



GB 18030-2022 中所示字形仅作为该图形字符的参考字形。汉字及其他图形字符的具体字形应符合相关字型国家标准的规定。

下面举例列出本标准图 C.1 中部分字符的字形、代码及其对应的 GB/T 13000 代码，以利读者理解。

8139

	30	31	32	33	34	35	36	37	38	39
EE	33F7	33F8	33F9	33FA	33FB	33FC	33FD	33FE	33FF	止 3400
EF	𠂇 3401	𠂈 3402	𠂉 3403	𠂊 3404	𠂋 3405	𠂌 3406	𠂍 3407	𠂎 3408	𠂏 3409	𠂐 340A

上图是本标准图 C.1 的节选，其中各字符的字形、代码及其对应的 GB/T 13000 代码分别如下表所示：

字形	GB 18030 代码	GB/T 13000 代码	字形	GB 18030 代码	GB/T 13000 代码
(无字形)	8139EE30	33F7	𠂇	8139EF30	3401
(无字形)	8139EE31	33F8	𠂈	8139EF31	3402
(无字形)	8139EE32	33F9	𠂉	8139EF32	3403
(无字形)	8139EE33	33FA	𠂊	8139EF33	3404
(无字形)	8139EE34	33FB	𠂋	8139EF34	3405
(无字形)	8139EE35	33FC	𠂌	8139EF35	3406
(无字形)	8139EE36	33FD	𠂍	8139EF36	3407
(无字形)	8139EE37	33FE	𠂎	8139EF37	3408
(无字形)	8139EE38	33FF	𠂏	8139EF38	3409
止	8139EE39	3400	𠂐	8139EF39	340A

3.4 附录 D（资料性）部分字符和代码的说明

【标准条文】

D.1 调整过的 18 个 GB/T 13000 代码位置

GB 18030-2005 规定了 10 个竖排标点和 8 个汉字构件在 GB/T 13000 中的临时代码位置，这些代码位置在本文件中进行了调整，如表 D.1 所示。

表 D.1 调整后的代码对应关系（略）

D.2 在双字节区删除的 6 个汉字构件



6 个汉字构件在 GB 18030-2005 中各自占据 2 个代码位置，本文件当前版本保留各自 2 个代码位置中的 1 个，如表 D.2 所示。

表 D.2 代码位置和保留的字形（略）

D.3 在双字节区删除的 9 个 CJK 兼容汉字

本文件删除了 GB 18030-2005 所收录的来自 GB/T 13000 的 9 个 CJK 兼容汉字，如表 D.3 所示。

表 D.3 删除的 9 个 CJK 兼容汉字和它们的代码位置（略）

【条文解读】

3.4.1 调整过的18个GB/T 13000代码位置

表 D.1 列出了“10 个竖排标点和 8 个汉字构件”调整前后的代码位置。

例如，竖排问号“？”的码位进行了如下调整：

本标准及 GB 18030-2005 的码位	GB 18030-2005 对应的 GB/T 13000 码位	本标准对应的 GB/T 13000 码位	字形
A6DF	E793	FE16	？
84318332	FE16	E793	(无字形)

本标准代码 0xA6DF 对应的 GB/T 13000 代码改为 0xFE16；

本标准代码 0x84318332 对应的 GB/T 13000 代码改为 0xE793。

即本标准代码 0xA6DF 和 0x84318332 分别对应的 GB/T 13000 代码进行了交换。

3.4.2 在双字节区删除的6个汉字构件

D.2 列出的 6 个汉字构件“𠂇 𠂈 𠂉 𠂊 𠂋 𠂌”在本标准上一版本中有新旧两个编码位置。至本标准第三版起草时，经调查，未发现在旧码位使用这 6 个汉字构件的产品，因而标准起草组决定仅保留与 GB/T 13000 对应的新码位，删除旧码位。

3.4.3 在双字节区删除的9个CJK兼容汉字

关于 CJK 兼容汉字的介绍，见条文释义部分的第 7 章。GB 18030—2005 收录了 21 个来自 GB/T 13000 的 CJK 兼容汉字，其中 12 个最终确认应属非兼容

汉字，其代码保持不变。这 12 个汉字位于本标准双字节 4 区，字形和代码见本标准第 7 章的释义部分。其它 9 个兼容汉字，不宜在我国使用，因而删除。

3.5 附录 E（资料性）《通用规范汉字表》汉字的代码位置

【标准条文】

本文件收录了 2013 年国务院批准发布的《通用规范汉字表》中的全部 8105 个规范汉字，它们的代码位置和字形见表 E.1。表中汉字的排列顺序与《通用规范汉字表》的规范汉字顺序一致，序号 1~序号 3500 的汉字为一级汉字，序号 3501~序号 6500 的汉字为二级汉字，序号 6501~序号 8105 的汉字为三级汉字。

表 E.1 《通用规范汉字表》汉字的字形和代码位置（略）

【条文解读】

标准原文的意思表达较为明晰，在此不作进一步解释。原文表 E.1 内容的电子信息见本文件的电子附件 2：《<通用规范汉字表>汉字的 GB 18030-2022 与 UCS 代码映射表》。

第四章 常见标准使用问题解答

4.1 GB 18030 标准相关问题

4.1.1 什么是编码字符集？

编码字符集是编码字符的集合。编码字符则指字符及其编码表示。

例如：汉字“啊”在 GB 18030-2022 中的编码是 0xB0A1。其中“啊”为字符，0xB0A1 为该字符的编码表示。

4.1.2 什么是GB 18030标准？

GB 18030《信息技术 中文编码字符集》是我国自主研制的以汉字为主的、包含多种我国少数民族文字（如蒙古、藏、维吾尔、朝鲜、彝、傣文等）的超大型中文编码字符集国家标准。GB 18030 第一版为 2000 版，第二版为 2005 版，其最新版本第三版为 2022 版。

4.1.3 GB 18030收录汉字的数量和范围？

GB 18030-2022收录中、日、韩（CJK）统一汉字、兼容汉字以及康熙部首、补充CJK部首共计88115个，主要来自《新华字典》、《汉语大字典》、《汉语大词典》、《辞源》、《辞海》、《康熙字典》、《现代汉语通用字表》、《简化字总表》等我国常见辞书和有关汉字使用的规范性文件。

GB 18030-2022收录的汉字覆盖了当前我国在社会生活的各个领域以及公安、金融、国土、交通、出版等专业领域的绝大部分信息化用汉字。

4.1.4 GB 18030三个版本的区别和关系？

——2000 版（已废止）

GB 18030 第一版于 2000 年发布，标准名称为《信息技术 汉字编码字符集基本集的扩充》。该版本规定了常用非汉字符号和 27533 个汉字（包括部首、部件等）的编码。GB 18030-2000 是全文强制的。

——2005 版（已废止）

GB 18030 第二版于 2005 年发布,标准名称为《信息技术 中文编码字符集》,自发布之日起代替 GB 18030-2000。该版本为部分强制,其中强制性部分与 2000 版一致。在此基础上,2005 版和 2000 版的主要区别是增加规定了 42711 个汉字和多种我国少数民族文字的编码,这些是推荐性的。

——2022 版（现行版本）

GB 18030 第三版于 2022 年发布,自发布之日起代替 GB 18030-2005。该版本为全文强制。

与 2005 版相比,2022 版除增加了编码汉字的数量外,还增加了四种少数民族文字。其它重要修改包括:

- 增加“实现的级别”一章

在标准文本中增加“实现的级别”一章,以实现本标准中字汇(指字符的数量和编码范围)的大小为依据,将产品的实现级别分为 3 级。

- 增加关于《通用规范汉字表》的资料性附录

增加资料性附录,列出表中汉字的字形和代码位置,以方便厂商实现《通用规范汉字表》。

除上述三个版本之外,目前正在研究制定 GB 18030-2022 第 1 号修改单。修改单预计将在现行标准基础上再增加 9000 余个编码汉字,主要来源为已经收入国际标准 ISO/IEC 10646: 2020《信息技术 通用编码字符集》及其补篇但尚未收入 GB 18030-2022 的 CJK 统一汉字及其扩充汉字,和我国公安人口信息专用字库补充汉字。修改单正式实施后,现行版本标准内容将对应变更。

4.1.5 GB 18030与GB/T 2312、GBK、GB/T 13000、ISO/IEC 10646、Unicode Standard等标准和文件的关系?

- 与 GB/T 2312 的关系

GB/T 2312-1980《信息交换用汉字编码字符集 基本集》收录了 6763 个汉字,GB 18030 收录了 GB/T 2312 规定的全部字符。在系统实现上,采用 GB 18030 可以完全兼容 GB/T 2312。

同时具备中文信息处理和交换功能的信息技术产品应符合 GB 18030 标准，不符合上述条件的产品可采用 GB/T 2312。（见本章“8、GB 18030 适用于什么产品？”）

● 与 GBK 的关系

GBK 于 1995 年发布，全名是《汉字扩展内码规范（GBK）1.0 版》，目前已废止。GBK 收录汉字和部件 20000 余个，兼容 GB/T 2312。

GB 18030 是在 GBK 基础上制定的，完全兼容 GBK。自 GB 18030 – 2000 发布之日起，GBK 即已废止。

● 与 GB/T 13000 和国际标准 ISO/IEC 10646 的关系

GB/T 13000《信息技术 通用多八位编码字符集（UCS）》等同采用国际标准 ISO/IEC 10646。国际标准 ISO/IEC 10646 对全世界古今文字统一编码，解决多文种统一处理的问题。我国自 1980 年代至今，积极参与 ISO/IEC 10646 的制定，是该标准工作组（ISO/IEC JTC1/SC2/WG2）中最活跃的成员之一。

GB 18030 采用的编码体系与 GB/T 13000 不同。GB 18030 可以完全兼容 GB/T 2312 和 GBK，而 GB/T 13000 不能兼容 GB/T 2312 和 GBK。GB 18030 规定的字符及其编码可与 GB/T 13000 一一对应，字符的编码可通过映射表⁶在两种编码体系之间自由转换。

GB 18030 的设计目的是解决我国文字信息处理问题，仅收录了汉字、我国少数民族现行文字以及部分在我国常用或常见的文字、符号。而 GB/T 13000（ISO/IEC 10646）的设计目的是解决全世界古今文字的信息处理问题，因而其收录文字、符号的种类和数量多于 GB 18030。

● 与产业规范 Unicode Standard 的关系

产业规范 Unicode Standard 由 Unicode 组织（Unicode Consortium）提出，其编码结构、收录字符与国际标准 ISO/IEC 10646 保持一致，不同之处在于 Unicode Standard 对各文种的编码实现作了更详细的规定。

虽然从技术角度看，产业规范 Unicode Standard 与国际标准 ISO/IEC 10646

⁶ 请见本文件附件 1：《GB 18030-2022 与 UCS 的代码映射表》。

基本一致,但由于 ISO 世界标准服务网公布的国际标准化机构或组织中并不包含 Unicode 组织,因此 Unicode Standard 并非国际标准。

4.2 政策相关问题

4.2.1 不符合GB 18030的产品如何处理?

根据中华人民共和国标准化法第二条,强制性标准必须执行;第二十五条,不符合强制性标准的产品、服务,不得生产、销售、进口或者提供。

标准发布后,企业应针对标准要求对产品开展技术改造,并在过渡期内具备生产(或提供)满足新标准的产品(或服务)的能力。

标准正式实施时,所有在中国市场上销售的产品均应符合新版标准,不符合新版标准的产品或服务不应再进行销售或提供。

4.2.2 在标准发布前已经进入流通领域的产品是否需要符合标准?

GB 18030-2022 于 2022 年 7 月 19 日发布,自 2023 年 8 月 1 日起实施。从国家标准发布到实施之间的时间段称为国家标准过渡期。之所以设置过渡期,既为企业开展技术改造、顺利过渡到生产(或提供)满足新标准的产品(或服务)留出时间,也是为消化已经上市的产品留出时间。当过渡期结束,标准正式实施时,所有在中国市场上销售的产品均应符合新版标准。

对于提供公共服务的机构应对其正在使用的系统进行升级处理,以便可以符合 GB 18030 相应等级的要求。

4.3 产品实现应关注的问题

4.3.1 GB 18030适用于什么产品?

简单地说:GB 18030标准适用于同时具备中文信息处理和交换功能的信息技术产品。

下面列出了6类具代表性的产品和服务:

——**具有中文信息处理功能的软件产品。**如各类操作系统、字库、输入法、文字处理、语音处理、文档编辑、机器翻译、汉字简繁转换等软件产品。

注:中文输入法产品,应包括但不限于键盘、手写、语音输入法产品。既包括作为单独商品出售的输入法产品,也包括作为商品组成部分的输入法产品,如 MS Windows、Mac OS、

Linux 等桌面操作系统，以及 Android、iOS 等移动操作系统自带的中文输入软件。

——**具有中文信息处理功能的硬件产品。**即安装前述软件产品的计算机、打印机、电信终端设备（如移动通信终端）、电子书阅读器、学习机等硬件产品。

——**数字出版物。**如数字化教材、图书、报纸、期刊、课件、游戏、地图，以及以网络形式出版的文字作品等。

——**数字传媒。**如互联网媒体，包括公共信息类网站、微信公众号等。

——**其他对语言文字进行信息化处理或内容编辑的信息技术产品。**如各类应用软件。

——**具有中文信息处理需要的行业应用。**如户籍、金融、交通、医疗、教育、供水、供电以及其它公共服务等。

4.3.2 如何理解GB 18030对产品的强制性要求？

GB 18030-2022 对产品的强制性要求分为两个层面，即编码层面和字汇层面。

——编码层面：

实现 GB 18030 相应级别的产品，接收到该实现级别内规定的编码数据作为输入时，产品各项功能均应正常，不应出现崩溃、报错、得到错误输出结果等现象，这里所指的输出，包括屏幕显示、打印输出、保存、发送等；

——字汇层面：

实现 GB 18030 相应级别的产品，应能保证在产品的交互界面上正确输出该实现级别内规定的所有图形字符，这里所指的输出，包括屏幕显示、打印输出、保存、发送等。

需要注意的是，GB 18030 是编码标准，而不是字型标准。GB 18030-2022 主要规定了汉字等中文图形字符的编码，而标准中所示字形仅作为该图形字符的参考字形。汉字及其他图形字符的具体字形应符合相关字型国家标准的规定。

4.3.3 GB 18030的3个实现级别分别对应什么产品？

● 实现级别 1：

所有适用于 GB 18030-2022 的产品和服务，最低应满足实现级别 1 的要求。可参见本章问题 8 中列举的 6 类产品和服务。

● 实现级别 2：

除标准中明确规定需达到实现级别 2 的系统软件和支撑软件产品外，问题 9 中提到的学习机、数字化教材、课件等与中文教学相关的产品也建议符合实现级别 2。因为这些产品所处理的汉字，往往涉及中国历史、地理、语文教学中不可避免出现的人名、地名、少数民族语言或外国语言音译汉字，以及科学教学中不可避免出现的近、现代创造的科技用汉字。这类产品如果仅支持本标准实现级别 1 部分的汉字，显然无法满足使用需要，以至少满足实现级别 2 的要求为宜。

● 实现级别 3:

问题 9 列举的 6 类产品和服务中，第六类主要涉及政务服务和公共服务等行业应用，此类服务因为涉及实名注册、验证等功能，对人名、地名用字处理能力有较高的需求，因而应满足实现级别 3 的要求。

4.3.4 符合 GB 18030-2005 强制部分要求的产品，是否可以认为符合 GB 18030-2022 的实现级别 1?

符合 GB 18030-2005 强制部分要求的产品经字库升级后可以符合 GB 18030-2022 的实现级别 1。相较 GB 18030-2005 强制部分，2022 版标准实现级别 1 中增加了 66 个 CJK 统一汉字，删除了 6 个汉字构件、9 个 CJK 兼容汉字，调整了 18 个字符的 GB/T 13000 代码位置。

4.3.5 产品达到什么要求才算符合 GB 18030?

——同时达到以下两个要求的系统，称为符合 GB 18030-2022 实现级别 1:

系统可以正确输入、输出、处理 GB 18030-2022 实现级别 1 规定的全部汉字字符；

系统可以正确识别 GB 18030-2022 实现级别 1 规定的全部汉字字符对应的编码。

——同时达到以下两个要求的系统，称为符合 GB 18030-2022 实现级别 2:

系统可以正确输入、输出、处理 GB 18030-2022 实现级别 2 规定的全部汉字字符；

系统可以正确识别 GB 18030-2022 实现级别 2 规定的全部汉字字符对应的编码。

——同时达到以下两个要求的系统，称为符合 GB 18030-2022 实现级别 3:

系统可以正确输入、输出、处理 GB 18030-2022 实现级别 3 规定的全部汉字字符及康熙部首；

系统可以正确识别 GB 18030-2022 实现级别 3 规定的全部汉字字符及康熙部首对应的编码。

4.4 消费者使用中文信息技术产品时的常见问题

4.4.1 消费者如何判断产品是否符合GB 18030标准？

- 询问产品生产商，该产品是否符合 GB18030 标准；
- 查阅产品文档，是否有该产品符合 GB18030 标准的相关声明；
- 到国家市场监督管理总局全国认证认可信息公共服务平台上查询，该产品是否通过 GB 18030 认证，网址如下：
<http://cx.cnca.cn/CertECloud/index/index/page>；
- 联系中国电子技术标准化研究院赛西实验室，咨询该产品是否通过了 GB 18030 标准符合性检测。咨询电话：010-64102066。

4.4.2 作为用户，是否使用了符合GB 18030标准的产品就能处理全部生僻字？

从产品方面来看，实现生僻字的显示、处理、输出等功能，需要您使用的字库、操作系统、应用软件、打印机等软硬件产品均能支持 GB 18030 标准，且建议应达到标准的实现级别 3。

从 GB 18030 标准本身来看，尽管 GB 18030-2022 已收录了 8 万余个汉字，但仍有少部分生僻字未进入标准，如公安人口信息专用字库中的一些人名、地名生僻字，以及尚未编码的古籍、科技用字。因此，可能存在 GB 18030 标准无法涵盖您希望处理的生僻字的情况。

生僻字的搜集整理是一项长期的系统性工程。在未来的一段时间内，GB 18030 标准起草组还将持续对新发现的生僻字进行审查和增补，以期更高质量地满足社会各界对于生僻字的信息化处理需求。

4.5 检测及认证相关问题

4.5.1 标准符合性检测和认证的定义和区别？

标准符合性检测，是以标准中的要求性条款为依据，对产品的相应指标进行检测，以确定产品是否符合标准要求的检测工作。标准符合性检测的对象，是标准适用范围内的产品。在检测过程中关注的测试项是标准中规定的内容；标准中未规定的项目不在标准符合性检测范围之内。

认证是指由国家认可的认证机构证明一个组织的产品、服务、管理体系符合相关标准、技术规范或其强制性要求的合格评定活动。针对 GB 18030 开展的认证是一种产品认证，即由认证机构证明一个组织的产品符合 GB 18030 的相关要求。

二者最重要的区别在于：

- 检测合格后，由检测机构出具检测报告，给出被测样品是否符合 GB 18030 的断语。该结论仅对被测样品本身负责；
- 认证合格后，由认证机构颁发认证证书及认证标识，其结论适用于与被测样品具有相同名称及版本号的所有产品，同时可在其产品上使用认证标识（类似于一个 logo）。GB 18030 认证过程中的一个重要环节，就是标准符合性检测。

4.5.2 GB 18030的检测和认证机构

检测机构：中国电子技术标准化研究院赛西实验室（中国合格评定国家认可委员会实验室认可证书注册编号：CNAS L0357）

联系人：王欣 13911782773 何正安 18610355189

认证机构：北京赛西认证有限责任公司（机构批准号：CNCA-R-2002-011）

联系人：崔莉 010-64102692 / 13552906796

4.5.3 什么样的产品适用于GB 18030标准符合性检测？

GB 18030-2022 标准适用范围内的所有产品都适用于 GB 18030 标准符合性检测。

除不与信息技术产品和设备进行信息交换的设备，以及仅提供中文提示信息的设备外，凡具有中文信息处理和交换功能的设备都应符合 GB 18030-2022 的要求。

4.5.4 什么样的产品不适用于GB 18030标准符合性检测？

符合以下任一条件的产品不适用于 GB 18030 标准符合性检测：

- 不进行中文信息处理的产品；
- 不与其他信息处理系统进行中文信息交换的产品。

4.5.5 需要对产品中哪类功能进行标准符合性检测？

产品中，与中文图形字符信息化处理及交换相关的功能需要进行 GB 18030 标准符合性检测。

不处理中文图形字符的功能不在测试范围内。如绘制图形、拍摄照片等功能。（若有绘制图形时加入文字或处理照片时加入文字的功能，则在测试范围内。）

不进行中文图形字符信息交换的功能不在测试范围内。如只能显示固定内容的办公设备和家用电器产品等。

这里的中文图形字符指 GB 18030 标准中规定的字符，而非仅指汉字及各少数民族文字。

4.5.6 检测范围内的哪些功能应符合标准要求？

所有可由最终用户自定义数据的功能，必须能正确处理按照 GB 18030 规定进行编码的数据。例如：

- 在数据库或文档中编辑数据；
- 自定义用户文件夹/文件/路径名称；
- 自定义用户帐户的显示名称；
- 自定义用户数据中用于数据分类、检索、推荐的关键词等。

下述功能现阶段暂不做要求：

- 解析网址；
- 解析电子邮件地址；
- 解析域名；
- 账户名称及密码（实名制账户除外）；
- 系统为后端使用而创建的字符串；
- 用于日志或管理员/开发者使用的以诊断为目的的数据；

- 命令名；
- 变量、ID、标记、对象、表等在编程语言或脚本中使用的名称；
- 在命令行状态下输入指令等。

4.5.7 企业对产品进行自测时应注意什么？

企业自测时，作为测试输入的数据应包含标准中相应级别范围内的所有编码。

在进行标准符合性测试时，鉴于字符及编码数量过于庞大，作为测试输入的数据可以采用一定的策略从标准中相应级别范围内的所有编码中抽取，同时应给出抽取原则和方法，并记录抽取的所有编码。抽取时还应注意：

- 应覆盖相应级别范围内的所有编码区域；
- 每个编码区域均应抽取足够的数量；
- 对于易出错的区域在抽取策略上应有侧重。

4.5.8 测试中有哪些常见的检出问题？

——占用用户自定义区

产品检测的统计结果显示，检出问题最多的是产品占用用户自定义区内的编码。

GB 18030 标准在双字节编码空间内规定了 3 个用户自定义区，这些编码空间是留给产品的最终用户使用的，禁止开发商在开发自己的产品时占用。

——双字节 4 区末尾

在产品检测中经常检测出产品在识别双字节 4 区末尾的某些编码时，无法正确显示相应的字符。造成该问题的原因，是这些字符的编码在国际标准中进行了重新分配，而产品没能将新分配的国际标准编码正确映射至 GB 18030 编码。

——实现级别 3 的产品未支持康熙部首

在检测中发现，较多标称实现级别 3 的字库产品虽满足支持标准内全部汉字的要求，却忽略了实现级别 3 中对支持康熙部首的要求。



第五章 标准相关参考资料选编

5.1 编码字符集标准的演进

受限于编码空间,早期的编码字符集只能编码单一文种或同时编码少数几个文种,如 GB 1988《信息技术 信息交换用七位编码字符集》(idt ISO/IEC 646,业界亦称 ASCII 码)、GB 2312《信息交换用汉字编码字符集 基本集》以及 GB 12052《信息交换用朝鲜文编码字符集》等⁷。

20 世纪晚期,多文种处理技术兴起。针对这种情况,1991 年,技术联盟 Unicode 组织(Unicode Consortium)发布了 Unicode 规范(Unicode Standard v1.0),而国际标准化组织/国际电工委员会 第 1 联合技术委员会(ISO/IEC JTC1)下属字符集与编码分技术委员会(SC2)也提出了国际标准 ISO/IEC 10646《信息技术 通用多八位编码字符集(UCS)》⁸计划。在此期间,ISO/IEC JTC1/SC2 与 Unicode 组织达成一致意见,两个标准采用共同的结构和字汇,并在发布进度上尽力一致。1993 年国际标准 ISO/IEC 10646 第一版发布,1997 年 Unicode 规范第二版(v2.0)发布,从此二者同步研制,并行不悖。我国积极参与国际标准 ISO/IEC 10646 的工作,于 1993 年同步发布了 GB 13000.1,等同采用国际标准 ISO/IEC 10646-1:1993。

1995 年,由于 GB 2312 收录的 6763 个汉字已不能满足当时信息化应用的需要,原国家技术监督局标准化司和原电子工业部科技与质量监督司联合发布了由全国信息技术标准化技术委员会组织制订的技术规范指导性文件《汉字扩展内码规范(GBK)1.0 版》。GBK 在编码方面采用当时汉字信息处理产品通用的内码方案,兼容 GB 2312;在字汇方面覆盖 GB 13000.1 的全部 20902 个汉字,并提供代码映射表。此外,还收录了此前不曾被国家标准或国际标准收录的 80 个汉字和部首/构件,13 个汉字结构符,以及它们在 GB 13000.1 中的临时代码。

⁷ 强制性国家标准 GB 1988、GB 2312、GB 12052 以及随后提到的 GB 13000 已于 2017 年转为推荐性标准。

⁸ 自 2012 年版开始,国际标准 ISO/IEC 10646 的名称不再保留“多八位”一词,因而未来 GB/T 13000 修订后的名称也不再保留“多八位”一词。

1990 年代末，GBK 方案已广泛使用。但由于 GBK 不是正式标准，在实施和市场监管方面有着固有缺陷，经主管部门协商，决定以 GBK 为基础制定强制性国家标准

2000 年，GB 18030 第一版，即 GB 18030—2000《信息技术 信息交换用汉字编码字符集 基本集的扩充》发布并立即实施。GB 18030—2000 收录汉字 27533 个，覆盖国际标准 ISO/IEC 10646: 2000 基本多文种平面（BMP）的全部汉字。GB 18030—2000 发布会由原信息产业部和原国家技术监督局共同主持，在新华社新闻发布大厅同时面向 30 多家新闻媒体召开，并开创了我国国家标准当天发布当天实施的先例。

2005 年，GB 18030 的第二版，即 GB 18030—2005《信息技术 中文编码字符集》正式发布。与第一版相比，第二版除补充了国际标准 ISO/IEC 10646: 2000 辅助多文种平面（SMP）的 42711 个汉字外，还增加了对我国部分少数民族文字字符编码的规定，包括蒙古文(含传统蒙古文、满文、锡伯文、托忒文和阿礼嘎礼文)、藏文、维哈柯文（包括维吾尔文、哈萨克文、柯尔克孜文）、朝鲜文、彝文和德宏傣文。GB 18030-2005 的名称从《信息技术 信息交换用汉字编码字符集 基本集的扩充》改为《信息技术 中文编码字符集》，充分反映了这一变化。

2022 年，GB 18030 的第三版，即 GB 18030—2022《信息技术 中文编码字符集》正式发布。其收录的中国文字有：

- 87887 个汉字（87875 个 CJK 统一汉字和 12 个 CJK 兼容汉字）；
- 228 个汉字部首（214 个康熙字典部首和 14 个补充 CJK 部首）；
- 10 种少数民族现行文字，即：蒙古文（包括传统蒙古文、满文、锡伯文、托忒文、阿礼嘎礼字）、藏文、维哈柯文（包括维吾尔文、哈萨克文、柯尔克孜文）、朝鲜文、规范彝文、滇东北苗文、傈僳文、西双版纳新傣文、西双版纳老傣文、德宏傣文。

GB 18030 作为收录汉字数量最多的中文编码字符集标准，自发布以来对我国中文信息技术产品市场产生了深刻而有益的影响，为我国大规模中文信息化建设奠定了坚实的标准化基础。



5.2 汉字的排序

GB 18030 所收汉字的排序在不同分区内遵循不同的规则。

5.2.1 双字节2区汉字的分级和排序

双字节 2 区编码的全部汉字来自于 GB/T 2312-1980。在本标准中，仍然按照 GB/T 2312-1980 的汉字排列顺序，进行排列。这里引用 GB/T 2312—1980 附录 2 《汉字的分级与排列》的部分文字如下，以便说明。

1. 汉字的分级

本标准收汉字 6763 个，分成两级。第一级汉字 3755 个，置于 16 区至 55 区；第二级汉字 3008 个，置于 56 区至 87 区。

.....

2. 汉字的排列

第一级汉字按汉语拼音字母顺序排列，同音字以笔形顺序横（一）、直（丨）、撇（丿）、点（丶）折（乙，包括乚等笔形）为序。起笔相同按第二笔，依次类推。

第二级汉字按部首排列，本标准采用的部首与一般字典用的部首基本相同，略有改并。部首次序及部首字按画数排列，同画数的字以笔形顺序横（一）、直（丨）、撇（丿）、点（丶）折（乙，包括乚等笔形）为序。起笔相同按第二笔，依次类推。

注：GB/T 2312 的 16 区至 55 区，即 GB 18030 双字节 2 区的表 B0 至表 D7；GB/T 2312 的 56 区至 87 区，即 GB 18030 双字节 2 区的表 D1 至表 F7。

5.2.2 双字节3区、4区以及四字节区汉字的排序

在双字节 3 区、4 区以及四字节区编码的汉字，全部来自国际标准 ISO/IEC 10646（idt GB/T 13000）。在本标准中，仍然按照 GB/T 13000 的汉字排列顺序，进行排列。

GB/T 13000 所收汉字，称为 CJK 统一汉字。其排序的基本规则为：全部汉字按照《康熙字典》部首归部；部首顺序按照《康熙字典》的规定；相同部首的汉字按照部首外笔画数增序排列，同画数的字以笔形顺序横（一）、直（丨）、撇（丿）、点（丶）折（乙，包括乚等笔形）为序。起笔相同按第二笔，依次类推。这里引用 GB/T 13000-2010 附录 S（资料性附录）《CJK 汉字的认同和排

序规则》的部分文字如下，以便说明。

S.2.1 排序的范围

第 33 章的 CJK 统一汉字代码表按以下字典先后顺序排列：

优先级	字典	版本
1	Kangxi Dictionary 《康熙字典》	北京第 7 版
2	Daikanwa Jiten 《大汉和辞典》	第 9 版
3	Hanyu Dazidian 《汉语大字典》	第 1 版
4	Daejaweon 《大字源》	第 1 版

这些字典是按上表指定的先后顺序使用的，1 级为最高级。已在高优先级字典中收录的汉字，不必再于低优先级的字典中查检。

S.2.2 排序规则

S.2.2.1 上述字典已收录的汉字

a)《康熙字典》已收录的汉字，则按《康熙字典》的顺序将其置于代码表中。

b)《康熙字典》未收录而《大汉和辞典》收录的汉字，则将其置于该字所属的部首—笔画的最后，并紧随前一个同时出现在《康熙字典》和《大汉和辞典》的汉字之后。

c)《康熙字典》和《大汉和辞典》均未收录的汉字，则按以上规则查检《汉语大字典》和《大字源》。

注：部首简化的汉字，置于该字对应繁体所属《康熙字典》相应部首的最后，并按笔画顺序排列。（此注由本国家标准起草者增补。）

S.2.2.2 上述字典均未收录的汉字

上述四部字典均未收录的汉字，置于该字所属部首—笔画的最后（上述四部字典所收录的汉字之后），并按同样的部首—笔画顺序排列。



5.3 中国民族古文字

本标准未收录中国民族古文字。

中国民族古文字是指我国古代历史上少数民族创造或从域外传入的记录中国境内民族语言的文字。这些文字曾在特定的历史时期使用，有些因使用这些文字的民族群体的分化融合或语言转用等原因而停止使用并逐渐消失，成为“死文字”；有些经过改进仍然为本民族使用，记录其现代语言。一般来说，民族文字的古和今以辛亥革命（1911年）为界，辛亥革命以前中国境内创制或流行的各种记录民族语的文字均可称作民族古文字。民族古文字的产生和使用年代最早可追溯到公元2—3世纪，如大约于公元2世纪传入于阗（今新疆和田）、鄯善（今新疆若羌）一带的佉卢文。而创制于16世纪的清代通用文字满文则一直沿用至近现代。惟僛僛族哇忍波创制于1923—1927年的僛僛音节文字因其形态原始，一直以来也被归入民族古文字当中。中国是世界上文字种类最丰富的国家，从远古到20世纪初，中华大地上曾经出现过近40种少数民族文字，他们和汉字一起共同形成了我国绵延不断的五千年文明。

我国古代民族文字的产生来源异常多样，除了仿汉字的方块壮字，变汉字的契丹大字、契丹小字、女真文、西夏文之外，有些出自域外非常古老的阿拉美字母，如佉卢文、焉耆-龟兹文、于阗文、粟特文、突厥文。而回鹘文又是在粟特文的基础上发展演变而成，从回鹘文进而衍生出来的有蒙古文、满文、锡伯文等。作为元朝国书的八思巴字因藏文而造，而藏文则出自印度的婆罗米字母。

按照文字的发生学关系，学界一般把古往今来中国境内出现和使用的民族古文字做如下分类：

1、阿拉美字母系文字

(1) 佉卢文

(2) 婆罗米字母文字：焉耆—龟兹文、于阗文；藏文、八思巴文；西双版纳老傣文、西双版纳新傣文、德宏傣文、傣绷文、金平傣文。

(3) 突厥文

(4) 粟特系文字：粟特文、回鹘文、蒙古文、托忒蒙古文、满文、锡伯文。

2、汉字系文字

仿汉字：方块白文、方块苗文、方块壮字、方块瑶文、方块布依文、方块侗

文、方块毛南字、方块哈尼文。

变汉字：契丹大字、契丹小字、女真文、西夏文、水书。

3、阿拉伯系文字

察合台文、维吾尔文、哈萨克文、柯尔克孜文、乌孜别克文、塔塔尔文。

4、自源文字

云贵川老彝文、纳西东巴文、哥巴文、尔苏沙巴文、傣傣音节文字。

5、拉丁字母文字

老傣傣文、滇东北老苗文、波拉字母傣傣文、撒拉佉文等。



5.4 CJK 统一汉字及其扩充的介绍

GB 18030《信息技术 中文编码字符集》收录的汉字，全部来自于 ISO/IEC 10646 (idt GB/T 13000)《信息技术 通用多八位编码字符集 (UCS)》中的 CJK 统一汉字及其扩充。在此对其进行简要介绍。

5.4.1 国际标准 ISO/IEC 10646 的介绍

国际标准 ISO/IEC 10646《信息技术 通用编码字符集 (UCS)》对全世界古今文字统一编码，以解决多文种统一处理的问题。与之对应的是我国国家标准 GB/T 13000。

除汉字外，国际标准 ISO/IEC 10646 收录了全部我国少数民族现行文字，如蒙古文、藏文、维哈柯文等。此外，还收了八思巴文、西夏文、契丹小字等我国古文字，以及阴阳、八卦、六十四卦、算筹、麻将、象棋等符号。

5.4.2 CJK 统一汉字的介绍

汉字的数量很大。不考虑字体差别（如楷体、宋体等），仅中国现存汉字就在 10 万字以上。又经过数千年使用和流传，汉字在不同国家和地区形成了不同的书写风格和造型差异，总量近百万。

显然，因为字形上的细微差异而对大量可以认同的汉字分别编码，在使用中不能带来明显好处，在某些应用场景中反而带来不便，而带给编码字符集标准研制和信息系统研发工作的负担是巨大的。1990 年代，计算机运算速度和存储能力仍然有限。国际标准 ISO/IEC 10646 中使用效率最高的双字节编码空间，即基本多文种平面 (Basic Multilingual Plane, BMP)，仅有 65536 个编码位置，能够分配给汉字的编码位置不足三分之一，远远不能满足对全部汉字编码的要求。因此，国际标准 ISO/IEC 10646《信息技术 通用编码字符集》的起草者决定对世界各国和地区使用的汉字统一编码。

由于中国（包括大陆和台湾地区）、日本、韩国最早参与汉字编码，所以国际标准 ISO/IEC 10646 把这些统一编码的汉字称为 CJK 统一汉字 (CJK UNIFIED IDEOGRAPHS)。所谓 CJK，即中国 (China，包括港澳台地区)、日本 (Japan)、韩国 (Korea)。CJK 统一汉字现在也包括了在越南 (Viet Nam) 和其它国家/地区 (如新加坡、马来西亚等) 使用的汉字，所以 CJK 统一汉字有时也被称为 CJKV 汉字，但在标准中仍保持 CJK 统一汉字的名称。

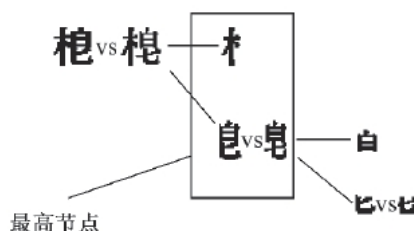
汉字编码工作由国际组织 ISO/IEC JTC1/SC2/WG2/IRG（Ideographic Research Group 汉字编码国际工作组）负责。有关 IRG 的情况，请参阅本章下一节“汉字编码国际工作组（IRG）”。

5.4.3 汉字的认同

为了决定哪些汉字可以使用同一编码，必须设计一套认同规则，在不影响使用的前提下忽略字形之间的细微差异，使得“相同”的汉字可以共用一个编码。所谓“相同”指的是抽象字形的相同，而不考虑其字音和字义。这套规则就是国际标准 ISO/IEC 10646 所附的汉字认同规则。基本规则如下：

- 1) 字形相近的汉字，比较其整体结构；
- 2) 如果整体结构相同，比较其细部结构，直至无可拆分；
- 3) 各级结构皆同，比较同级结构部件；
- 4) 各级部件也相同，则视为相同汉字。

下图摘自 GB/T 13000-2010（idt ISO/IEC 10646: 2003）《信息技术 通用多八位编码字符集（UCS）》的附录 S（资料性附录）《CJK 汉字的认同和排序规则》，以助理解。



国际标准 ISO/IEC 10646 的 2003 年版开始，汉字认同规则成为标准的规范性附录，附录顺序号为 S。目前该认同规则最新的中文翻译见 GB/T 13000-2010（idt ISO/IEC 10646: 2003）的附录 S（资料性附录）《CJK 汉字的认同和排序规则》。

虽然直到 2003 年，汉字认同规则的文本才正式收入国际标准 ISO/IEC 10646，但这套规则在 1990 年代即已实际运用，国际标准 ISO/IEC 10646 第一版（1993 年版）《信息技术 通用多八位编码字符集（UCS） 第一部分：体系结构和基本多文种平面》所收的 20902 个 CJK 统一汉字均已采用此汉字认同规则。

5.4.4 CJK 统一汉字的扩充

增补编码 CJK 统一汉字的工作持续至今未绝，称为对 CJK 统一汉字的扩充

(Extension)。

在 CJK 统一汉字字表中，编码序号自上向下顺序排列，相应的汉字字形也是自上向下顺序排列。因此，新增编码汉字的 CJK 统一汉字的扩充，又称为垂直扩充 (Vertical Extension，缩写为 VE)。又因为这种扩充是在现有 CJK 统一汉字字表之外进行，所以也称为外扩充。截至到 2022 年，国际标准 ISO/IEC 10646 中的 CJK 统一汉字 (包括扩充 A、B、C、D、E、F、G、H) 的数量达到 97036 个。其中，在基本多文种平面 (BMP) 收录了相对常用的 27581 个汉字，在辅助多文种平面 (Supplementary Multilingual Plane, SMP) 收录了 69455 个汉字。

与上面所提外扩充相对，在现有 CJK 统一汉字内部补充字形而不增加新的编码汉字的行为，称为内扩充 (Internal Extension，缩写为 IE；或 internal Supplement，缩写为 IS)。又因为这种行为导致字表中同一编码位置在水平方向的字形增加，所以又称为水平扩充或横扩充 (Horizontal Extension，缩写为 HE；或 Horizontal Supplement，缩写为 HS)。例如：

2541E
石 112.3



 韩国计划新增字形


第一次汉字内扩充由我国大陆提出。在 CJK 统一汉字的字表第一次发布时，所有字形按照中日韩等国家和地区分别列出。我国列下存在多处空缺，即此前我国未提出对该字编码，该字由其他国家或地区提出。为便于我国使用 ISO/IEC 10646 进行汉字信息交换，我国制定了 GB/T 16500-1998《信息交换用汉字编码字符集 第七辅助集》，所收汉字全部是上述空缺处的其他国家或地区提出编码的汉字。我国要求以 GB/T 16500-1998 为来源，在 CJK 统一汉字字表中我国一列填补字形空缺。其后多年，韩国等国家和地区也陆续提出了各自不同规模、不同编码位置的汉字内扩充。

5.4.5 CJK兼容汉字的介绍

按照 CJK 统一汉字认同规则，每个 CJK 兼容汉字都应与其某个 CJK 统一汉字认同，不能单独编码。但是，由于国家、民族、文化、历史等原因，某些国家标准/地区标准已经对同一个汉字重复编码。例如，韩国标准 KS C 5601-1987 (现已转为 KS X 1001:1998) 重复编码了 268 个分别具有多个读音汉字。按照 CJK

汉字认同规则，这 268 个汉字在 CJK 统一汉字区内只能分别编码一次，而不能重复编码。为了满足国际标准 ISO/IEC 10646 和 KS C 5601-1987 之间进行数据双向转换的要求，标准在 CJK 统一汉字区和 CJK 兼容汉字区(CJK COMPATIBILITY IDEOGRAPHHS)同时对这些汉字编码。

来自日本标准 JIS X 0213: 2000 的 59 个汉字，以及来自朝鲜、中国台湾地区等地标准的一些汉字，也作同样处理。

这里引用 GB/T 13000-2010 的第 22 章内容如下，以供参考。

22 兼容字符

GB 13000 中的兼容字符主要是为了与现存编码字符集进行双向代码转换，而不会丢失信息。

在 BMP 中，多数兼容字符分配在 F9、FA、FE 和 FF 行，以及 31 和 33 行。也有一些兼容字符分配在其他行。

注 1：BMP 平面 FA 行中的 12 个代码位置分配给 CJK 统一汉字，见第 A.1 章中 CJK UNIFIED IDEOGRAPHHS 汇集的定义。

在辅助表意文字平面(SIP)中，这些字符分配在 F8 到 FA 行。CJK 兼容汉字（CJK COMPATIBILITY IDEOGRAPHHS-2001 汇集的字符部分）应与某个 CJK 统一汉字（CJK UNIFIED IDEOGRAPHHS-2001 汇集的字符部分）认同，每条认同规则在本标准附录 S 中描述。

但是由于民族、文化和历史原因，某些国家和地区标准为这些字符分配了单独的代码位置，所以在本标准中它们也被作为独立的字符。

注 2：基于以上原因，兼容汉字只应该用于保证本标准与特定国家、地区或其他标准之间能够相互转换，其他用法都不赞成。

GB 18030 第三版删除了 2000 年版和 2005 年版收录的 9 个 CJK 兼容汉字。

这 9 个 CJK 兼容汉字分别具有明确对应的 CJK 统一汉字，对应关系如下：

GB 18030 第三版的码位	对应 GB/T 13000 兼容汉字的码位	对应 GB/T 13000 兼容汉字的字形	对应 GB/T 13000CJK 统一汉字的码位	对应 GB/T 13000CJK 统一汉字的
FD9C	F92C	郎	90CE	郎
FD9D	F979	凉	51C9	凉

FD9E	F995	季	5B63	季
FD9F	F9E7	裏	88CF	裏
FDA0	F9F1	隣	96A3	隣
FE40	FA0C	兀	5140	兀
FE41	FA0D	殼	55C0	殼
FE47	FA18	礼	793C	礼
FE49	FA20	蘊	8612	蘊

自 2000 年版开始，标准起草者就建议相关软件开发者和使用者首先选用与这 9 个 CJK 兼容汉字相对应的 CJK 统一汉字。自本标准第一次发布（2000 年版）至第三版（当前版本）已经 19 年，标准起草者观察其使用情况，认为已无必要在标准中继续保留这 9 个 CJK 兼容汉字。

因为双字节 4 区保留的 12 个 CJK 兼容汉字并无与之明确对应的 CJK 统一汉字，所以在产品实现时，这 12 个汉字应与本标准收录的 CJK 统一汉字一样处理。

（12 个兼容汉字，见本文件第 2 章第 7 节的说明。）

5.4.6 汉字部首和汉字排序

汉字在各个国家、地区的流传和使用过程中发生了演变，也包括了部首的增减和归部方法的变化。GB/T 13000 (idt ISO/IEC 10646) 中，CJK 统一汉字的排序采用部首、笔画数目规则。为统一规则，有关各方一致同意：凡需编码的汉字，一律归入《康熙字典》规定的 214 个部，其部首即所谓康熙部首 (KANGXI RADICALS)；汉字笔画数目的计算方式，以《康熙字典》规则为主；《康熙字典》选用北京中华书局影印本(1958 年 1 月第 1 版)。康熙部首在国际标准 ISO/IEC 10646 中的代码位置为 0x2F00 至 0x2FD5。

GB 18030 第三版收录了全部 214 个康熙部首，代码位置为 0x81398B32 至 0x8139A035。参见本文件第 2 章第 7 节“关于码位分配的补充说明”。

为了满足各方对康熙部首之外汉字部首的使用需要，GB/T 13000 又增收了中、日、韩等国家和地区常用的 115 个汉字部首，称为补充 CJK 补充部首 (CJK RADICALS SUPPLEMENT)。与康熙部首不同，CJK 补充部首不仅大量收录了我国简化汉字使用的部首（如“贝”“见”），还收录了仅能作为偏旁而非整字使用的现代部首（如“辶”“冫”）。CJK 补充部首在国际标准 ISO/IEC 10646 中的编码

位置为 0x2E80 至 0x2EF3（不包括 0x2E9A）。

GB 18030 第三版收录了其中 14 个，置于双字节 4 区的最后。参见本文件第 2 章第 7 节“关于码位分配的补充说明”。

5.5 汉字编码国际工作组（IRG）

IRG 全称为 Ideographic Research Group（曾用名 Ideographic Rapporteur Group），隶属于 ISO/IEC JTC1/SC2/WG2，专门负责国际标准 ISO/IEC 10646《信息技术 通用编码字符集（UCS）》中的中日韩统一汉字（CJK Unified Ideographs）的编码工作。IRG 没有正式的中文名称，一般翻译为汉字编码国际工作组。

汉字在中国、日本、韩国等东亚国家和地区广泛使用。虽然汉字在不同国家和地区的字形风格和使用方法有所不同，但存在着大量字形相同或基本相同的汉字。为了避免在标准中对来自不同国家或地区的“相同”汉字分别编码，浪费编码空间，增加信息处理系统的负担，ISO/IEC JTC1/SC2/WG2 决定成立 IRG，对汉字编码统一研究，统一制定。

IRG 的前身是通用中文代码国际联合会（ACCC，1984 年成立），中间经过中日韩联合研究组（CJK-JRG，1990 年由 ACCC 改组成立），至 1993 年正式改组为 IRG，运行至今。

目前，IRG 的成员包括中国、日本、韩国、越南、美国、英国、香港特别行政区、澳门特别行政区、台北电脑同业公会（TCA）和 Unicode 组织。新加坡、朝鲜、马来西亚也曾参加过 IRG 的活动。



附 件

本文件以电子附件形式提供《GB 18030-2022 与 UCS 的代码映射表》和《<通用规范汉字表>汉字的 GB 18030-2022 与 UCS 代码映射表》。

电子附件获取地址如下：

1. 全国信息技术标准化网：<http://www.nits.org.cn/index/page/article/1805>
2. “中文信息技术标准化”微信公众号

微信搜索“中文信息技术标准化”，或扫描以下二维码关注公众号。



进入发消息页面，下方菜单栏选择“实用工具”，获取下载地址。

