

ICS 35.040

L71



中华人民共和国国家标准

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

Information technology — Chinese ideograms coded character
set for information interchange — Extension for the basic set

(报批稿)

2000-03-17 发布

2000-03-17 实施

国家质量技术监督局 发布

前 言

本标准作为 GB 2311 体系的字符编码标准，规定了信息交换用的基本图形字符及其二进制编码的十六进制表示。

本标准适用于图形字符信息的处理、交换、存储、传输、显现、输入和输出。

本标准是对 GB 2312 的扩充。

本标准具体规定了图形字符的单字节编码和双字节编码，并对四字节编码体系结构做出了规定。

本标准的附录 A、附录 B、附录 C 和附录 D 是标准的附录。

本标准从生效之日期起，同时代替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以技监标函[1995]229 号文发布和实施的技术规范指导性文件《汉字内码扩展规范（GBK）》1.0 版。

本标准由中华人民共和国信息产业部提出。

本标准由信息产业部电子工业标准化研究所归口。

本标准起草单位：信息产业部电子工业标准化研究所、北京大学计算机技术研究所、北大方正集团、北京方正新天地信息网络科技有限责任公司、四通集团公司、中科院软件所、长城软件公司、四通利方公司、中软总公司、金山软件公司、联想公司。

本标准主要起草人：陈堃铎、黄疆、胡万进、张建国、陈壮。

中华人民共和国国家标准

信息技术 信息交换用汉字编码字符集 基本集的扩充

Information technology — Chinese ideograms

coded character set for information

GB 18030 — 2000

interchange — Extension for the basic set

1 范围

本标准作为 GB 2311 体系的编码字符标准，规定了信息交换用的图形字符及其二进制编码的十六进制表示。

本标准适用于图形字符信息的处理、交换、存储、传输、显现、输入和输出。

2 引用标准

下列标准所包含的条文，通过在本标准中引用而构成为本标准的条文。本标准出版时，所示版本均为有效。所有标准都会被修订，使用本标准的各方应探讨使用下列标准最新版本的可能性。

- GB 2311—1990 信息处理 七位和八位编码字符集 代码扩充技术 (eqv ISO 2022:1986)
- GB 2312—1980 信息交换用汉字编码字符集 基本集
- 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)
- GB 12345—1990 信息交换用汉字编码字符集 辅助集
- GB 13000.1—1993 信息技术 通用多八位编码字符集 (UCS) 第一部分：体系结构与基本多文种平面 (idt ISO/IEC 10646.1-1993)

3 原则

本标准向下与国家标准 GB 2312 信息处理交换码所对应的事实上的内码标准兼容。

本标准在字汇上支持 GB 13000.1 的全部中、日、韩 (CJK) 统一汉字字符和全部 CJK 统一汉字扩充 A 的字符。

4 定义

本标准采用下列定义。

4.1 字汇 repertoire

用编码字符集表示的一个指定的字符集合。

4.2 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。

4.3 编码字符 coded character

字符及其编码表示。

4.4 保留区 reserved zone

本标准中留作未来国际标准规定的区域。

5 字汇

本标准收录的字符分别以单字节、双字节和四字节编码。

5.1 单字节部分

本标准中，单字节的部分收录了 GB 11383 的 0x00 到 0x7F 全部 128 个字符及单字节编码的欧元符号。

5.2 双字节部分

本标准中，双字节的部分收录内容如下：

GB 13000.1 的全部 CJK 统一汉字字符。

GB 13000.1 的 CJK 兼容区挑选出来的 21 个汉字。

GB 13000.1 中收录而 GB 2312 未收录的我国台湾地区使用的图形字符 139 个。

GB 13000.1 收录的其它字符 31 个。

GB 2312 中的非汉字符号。

GB 12345 的竖排标点符号 19 个。

GB 2312 未收录的 10 个小写罗马数字。

GB 2312 未收录的带音调的汉语拼音字母 5 个以及 a 和 g 。

汉字数字“〇”。

表意文字描述符 13 个。

增补汉字和部首/构件 80 个。

双字节编码的欧元符号。

5.3 四字节部分

本标准的四字节的部份，收录了上述双字节字符之外的，包括 CJK 统一汉字扩充 A 在内的 GB 13000.1 中的全部字符。

6 总体结构

本标准中，采用单字节、双字节和四字节三种方式对字符编码。本标准中的任何一个字节均由八位二进制位串组成，任何一个八位的值均由 0x00 至 0xFF 的十六进制记数法表示。

单字节部分采用 GB 11383 的编码结构与规则，使用 0x00 至 0x80 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 11383 未采用的 0x30 到 0x39 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 0x81308130 到 0xFE39FE39。见表 1 及图 1。

表1 码位范围分配图

字节数	码位空间				码位数目
单字节	0x00~0x80				129个码位
双字节	第一字节		第二字节		23940个码位
	0x81~0xFE		0x40~0x7E, 0x80~0xFE		
四字节	第一字节	第二字节	第三字节	第四字节	1587600个码位。
	0x81~0xFE	0x30~0x39	0x81~0xFE	0x30~0x39	

四字节字符的编码自第四个字节开始，编码码位为 0x30 至 0x39；其次是第三个字节，编码码位为 0x81 至 0xFE；再次是第二个字节，编码码位为 0x30 至 0x39；最后是第一个字节，编码码位为 0x81 至 0xFE。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。

注：本标准中，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。

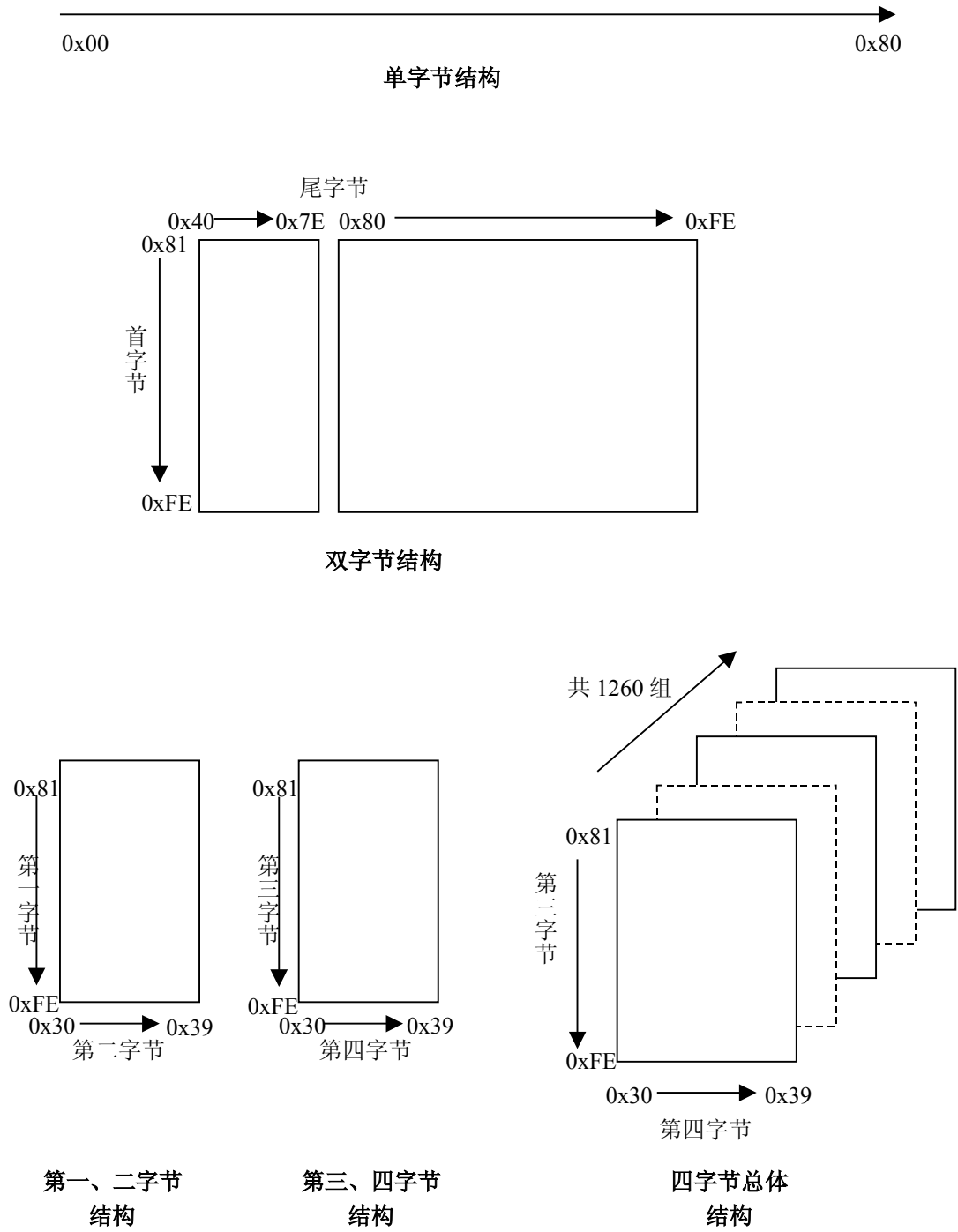


图 1 总体结构图

7 字符的排列顺序

7.1 单字节部分字符的排列顺序

本标准中单字节部分所有字符按照 GB 11383 中相应字符的顺序排列，单字节编码的欧元符号置于 0x80 位置，对应于 GB 13000.1 的 0x20AC 位置。见图 2。

7.2 双字节部分字符的排列顺序

本标准双字节部分的字符排列顺序见附录 A。

7.3 四字节部分字符的排列顺序

自 0x81308130 至 0x8439FE39 共 50400 个码位，对应本标准双字节部分未包括的所有 GB 13000.1 的字符，按照 GB 13000.1 相应字符的顺序排列，剩余码位保留。

自 0x85308130 至 0x8539FE39 共 12600 个码位，为本标准的保留区，留待未来字符扩展使用。

自 0x86308130 至 0x8F39FE39 共 126000 个码位，为本标准的保留区，留待未来汉字字符扩展使用。

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

自 0xE4308130 至 0xFC39FE39 共 315000 个码位，为本标准的保留区，留待未来标准扩展使用。

自 0xFD308130 至 0xFE39FE39 共 25200 个码位，为用户自定义区。

8 码位分配

8.1 单字节部分的码位分配

本标准中，单字节的部分的码位分配见 GB 11383。单字节编码欧元符号置于 0x80 位置。见图 2。

				b8	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
				b7	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
b4	b3	b2	b1	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	
0	0	0	0			SP	0	@	P	`	p	€								
0	0	0	1			!	1	A	Q	a	q									
0	0	1	0			"	2	B	R	b	r									
0	0	1	1			#	3	C	S	c	s									
0	1	0	0			¥	4	D	T	d	t									
0	1	0	1			%	5	E	U	e	u									
0	1	1	0			&	6	F	V	f	v									
0	1	1	1			'	7	G	W	g	w									
1	0	0	0			(8	H	X	h	x									
1	0	0	1)	9	I	Y	i	y									
1	0	1	0			*	:	J	Z	j	z									
1	0	1	1		ESC	+	;	K	[k	{									
1	1	0	0			'	<	L	\	l										
1	1	0	1			-	=	M]	m	}									
1	1	1	0			.	>	N	^	n	~									
1	1	1	1			/	?	O	_	o	DEL									

图2 单字节区码位图

8.2 双字节部分的码位分配

本标准中，双字节的部分的码位安排分为 0x8140 至 0xFE7E 和 0x8180 至 0xFEFE 两部分，共 23940 个码位。见图 3 及表 2。

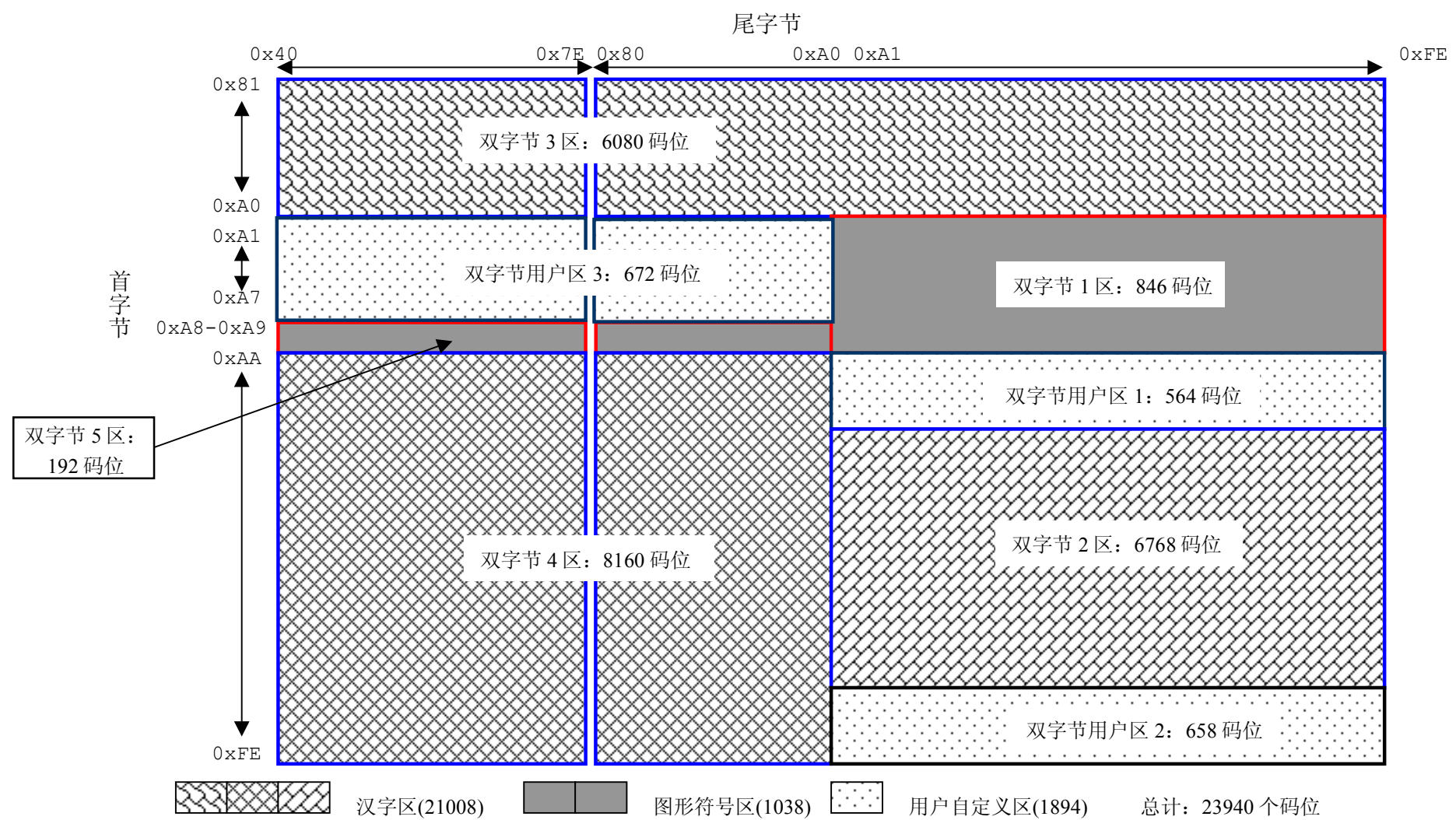


图3 双字节部分编码空间结构图

表2 双字节部分的码位安排

类别	区名	码位范围	码位数	字符数	字符类型
符号区	双字节1区	A1A1—A9FE	846	718	图形符号
	双字节5区	A840—A9A0	192	166	图形符号
汉字区	双字节2区	B0A1—F7FE	6768	6763	汉字
	双字节3区	8140—A0FE	6080	6080	汉字
	双字节4区	AA40—FEA0	8160	8160	汉字
用户自定义区	双字节用户区1	AAA1—AFFE	564		
	双字节用户区2	F8A1—FEFE	658		
	双字节用户区3	A140—A7A0	672		

本标准中，双字节部分的汉字区（即双字节2、3、4区）里，CJK统一汉字在前，增补汉字在后。其中GB 2312的编码汉字安排在双字节2区。GB 13000.1中选出的21个CJK兼容汉字编码在双字节4区的0xFD9C到0xFDA0和0xFE40到0xFE4F。增补汉字和部首/构件80个编码在双字节4区。

GB 13000.1中收录而GB 2312未收录的我国台湾地区使用的图形字符139个、汉字数字“〇”和表意文字描述符13个编码在双字节5区。

GB 2312中的非汉字符号、GB 2312未收录的带音调的汉语拼音字母5个以及 α 和 g 、GB 2312未收录的小写罗马数字10个、GB 12345的竖排标点符号19个和双字节编码的欧元符号（码位为0xA2E3）编码在双字节1区。

8.3 四字节部分的码位分配

四字节部分的码位分配见7.3条。

附录 A
(标准的附录)
双字节字符表

A1 表的内容

本表给出本标准双字节部分的全部 GB 13000.1 代码及其相应字形。

A2 表的说明

示例如下：

双字节 3 区

第一字节	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
6	𠂧 4E63	𠂨 4E64	𠂩 4E65	𠂪 4E67	𠂫 4E68	𠂬 4E6A	𠂭 4E6B	𠂮 4E6C	𠂯 4E6D	𠂰 4E6E	𠂱 4E6F	𠂲 4E72	𠂳 4E74	𠂴 4E75	𠂵 4E76	𠂶 4E77
7	𠂷 4E78	𠂸 4E79	𠂹 4E7A	𠂺 4E7B	𠂻 4E7C	𠂼 4E7D	𠂽 4E7F	𠂾 4E80	𠂿 4E81	𠃀 4E82	𠃁 4E83	𠃂 4E84	𠃃 4E85	𠃄 4E87	𠃅 4E8A	
8	𠃆 4E90	𠃇 4E96	𠃈 4E97	𠃉 4E99	𠃊 4E9C	𠃋 4E9D	𠃌 4E9E	𠃍 4EA3	𠃎 4EAA	𠃏 4EAF	𠃐 4EB0	𠃑 4EB1	𠃒 4EB4	𠃓 4EB6	𠃔 4EB7	𠃕 4EB8
9	𠃖 4EB9	𠃗 4EBC	𠃘 4EBD	𠃙 4EBE	𠃚 4EC8	𠃛 4ECC	𠃜 4ECF	𠃝 4ED0	𠃞 4ED2	𠃟 4EDA	𠃠 4EDB	𠃡 4EDC	𠃢 4EE0	𠃣 4EE2	𠃤 4EE6	𠃥 4EE7
A	𠃦 4EE9	𠃧 4EED	𠃨 4EEE	𠃩 4EEF	𠃪 4EF1	𠃫 4EF4	𠃬 4EF8	𠃭 4EF9	𠃮 4EFA	𠃯 4EFC	𠃰 4EFE	𠃱 4F00	𠃲 4F02	𠃳 4F03	𠃴 4F04	𠃵 4F05
B	𠃶 4F06	𠃷 4F07	𠃸 4F08	𠃹 4F0B	𠃺 4F0C	𠃻 4F12	𠃼 4F13	𠃽 4F14	𠃾 4F15	𠃿 4F16	𠄀 4F1C	𠄁 4F1D	𠄂 4F21	𠄃 4F23	𠄄 4F28	𠄅 4F29
C	𠄆 4F2C	𠄇 4F2D	𠄈 4F2E	𠄉 4F31	𠄊 4F33	𠄋 4F35	𠄌 4F37	𠄍 4F39	𠄎 4F3B	𠄏 4F3E	𠄐 4F3F	𠄑 4F40	𠄒 4F41	𠄓 4F42	𠄔 4F44	𠄕 4F45
D	𠄖 4F47	𠄗 4F48	𠄘 4F49	𠄙 4F4A	𠄚 4F4B	𠄛 4F4C	𠄜 4F52	𠄝 4F54	𠄞 4F56	𠄟 4F61	𠄠 4F62	𠄡 4F66	𠄢 4F68	𠄣 4F6A	𠄤 4F6B	𠄥 4F6D
E	𠄦 4F6E	𠄧 4F71	𠄨 4F72	𠄩 4F75	𠄪 4F77	𠄫 4F78	𠄬 4F79	𠄭 4F7A	𠄮 4F7D	𠄯 4F80	𠄰 4F81	𠄱 4F82	𠄲 4F85	𠄳 4F86	𠄴 4F87	𠄵 4F8A
F	𠄶 4F8C	𠄷 4F8E	𠄸 4F90	𠄹 4F92	𠄺 4F93	𠄻 4F95	𠄼 4F96	𠄽 4F98	𠄾 4F99	𠄿 4F9A	𠅀 4F9C	𠅁 4F9E	𠅂 4FA1	𠅃 4FA2		

第二字节高位
GB 13000.1 字形
GB 13000.1 代码

附 录 B
(标准的附录)
表意文字描述符

本表中收录了 GB 13000.1-1993 为收录的表意文字描述符 13 个。

编号	表意文字描述符	本标准的码位	GB13000.1的码位	功能描述
1		A989	E7E7	相似而不等，用于借某字表达外字。
2		A98A	E7E8	左右结构
3		A98B	E7E9	上下结构
4		A98C	E7EA	左中右结构
5		A98D	E7EB	上中下结构
6		A98E	E7EC	全包围结构
7		A98F	E7ED	向下包围结构
8		A990	E7EE	向上包围结构
9		A991	E7EF	向右包围结构
10		A992	E7F0	向右下包围结构
11		A993	E7F1	向左下包围结构
12		A994	E7F2	向右上包围结构
13		A995	E7F3	嵌套结构

附 录 C
(标准的附录)
追加的汉字及部首/构件

本表中收录了部分没有在 GB 13000.1-1993 中收录的汉字及部首/构件。

FE	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
5	E815	E816	E817	E818	E819	E81A	E81B	E81C	E81D	E81E	E81F	E820	E821	E822	E823	E824
6	E825	E826	E827	E828	E829	E82A	E82B	E82C	E82D	E82E	E82F	E830	E831	E832	E833	E834
7	E835	E836	E837	E838	E839	E83A	E83B	E83C	E83D	E83E	E83F	E840	E841	E842	E843	
8	E844	E845	E846	E847	E848	E849	E84A	E84B	E84C	E84D	E84E	E84F	E850	E851	E852	E853
9	E854	E855	E856	E857	E858	E859	E85A	E85B	E85C	E85D	E85E	E85F	E860	E861	E862	E863
A	E864															

目 次

前言

1 范围	1
2 引用标准.....	1
3 原则	1
4 定义	1
4.1 字汇	1
4.2 字符	1
4.3 编码字符	1
4.4 保留区	2
5 字汇	2
5.1 单字节部分.....	2
5.2 双字节部分.....	2
5.3 四字节部分.....	2
6 总体结构.....	2
7 字符的排列顺序	5
7.1 单字节部分字符的排列顺序	5
7.2 双字节部分字符的排列顺序	5
7.3 四字节部分字符的排列顺序	5
8 码位分配.....	5
8.1 单字节部分的码位分配.....	5
8.2 双字节部分的码位分配.....	6
8.3 四字节部分的码位分配.....	8
附录 A (标准的附录) 双字节字符表	9
附录 B (标准的附录) 表意文字描述符.....	91
附录 C (标准的附录) 追加的汉字与部首/构件	92
附录 D (标准的附录) 四字节字符表	93