

前 言

本标准等效采用国际标准 ISO/IEC 646:1991《信息技术——信息交换用七位编码字符集》。在 ISO/IEC 646 中基本代码表的基础上,并在其允许变动的范围内,结合我国的情况,补充了我国通常使用的人民币符号“元记号”,这与本标准第一版和第二版一致,该代码作为中国版本,于 1981 年提交 ISO 登记,并获得认可。该标准文本在内容和编写上与 ISO/IEC 646 文本有些差异,但在技术上是等效的。

代码表的内容是本标准的核心,中国方案原则上符合 ISO/IEC 646 基本代码表的规定,若与其国际参考方案比较,则只有代码表 2/4 位置上的图形符号形状不同,而其他图形符号完全一致。

本标准第二版和第三版的内容变更和主要差别见附录 B。

本标准引用的标准是 GB/T 5261—1994(idt ISO 6429:1988)版本,而 ISO/IEC 646(第三版)引用的是当时即将出版的 ISO/IEC 6429:1992。GB/T 5261 目前还未按 1992 年的 ISO/IEC 6429 修订。

本标准的附录 A 是标准的附录,附录 B 是提示的附录。

本标准由中华人民共和国电子工业部提出。

本标准由电子工业部标准化研究所归口。

本标准起草单位:电子工业部标准化研究所。

本标准主要起草人:向维良、王宝艾。

本标准首次发布(第一版)于 1980 年。

本标准第一次修订(第二版)于 1989 年。

ISO/IEC 前言

ISO(国际标准化组织)和 IEC(国际电工委员会)共同组成一个世界标准化专门系统。ISO 或 IEC 的国家成员体,通过涉及特殊技术活动领域的各个组织所建立的技术委员会来参与国际标准的开发。ISO 和 IEC 的技术委员会在共同感兴趣的领域内合作,与 ISO 和 IEC 有联络的其他官方和非官方国际性组织,也参与这项工作。

在信息技术领域内,ISO 和 IEC 已建立了一个联合技术委员会 ISO/IEC JTC1。由联合技术委员会采纳的国际标准草案需分发给各国家成员体表决。发布一项国际标准至少需要 75% 的国家成员体投票赞成。

国际标准 ISO/IEC 646 是由联合技术委员会 ISO/IEC JTC1“信息技术”制定的。

这个第三版对第二版进行了技术修订,取消并代替第二版(ISO 646:1983)。

附录 A 构成本标准整体的一部分,附录 B^{1]}和附录 C^{2]}只是提供信息。

采用说明:

- 1] ISO/IEC 646 的附录 B 是“从 ISO/IEC 646 中获得标准的指南”,本标准已由此得到了结果,勿需将国际标准的附录 B 的内容再放入本标准中。
- 2] ISO/IEC 646 的附录 C,在采用为本国家标准时改为附录 B。

中华人民共和国国家标准

信息技术 信息交换用七位编码字符集

GB/T 1988—1998
eqv ISO/IEC 646:1991
代替 GB/T 1988—1989

Information technology—7-bit Coded
character set for information interchange

1 范围

本标准规定了由 128 个字符(控制字符和图形字符,如:字母、数字及符号)组成的字符集以及它们的编码表示。

本标准规定了一个中国通用的代码表版本,也提供一个国际参考代码表版本。

本字符集主要用于数据处理系统与有关设备之间及数据通信系统内的信息交换,在确定本字符集时,还考虑了数据处理中所需的图形字符和控制功能。

当本字符集不能满足某些特殊应用场合需要时,它允许使用代码扩充控制字符,这些控制字符的使用方法在 GB/T 2311 中规定。

本标准中提到的控制字符的定义,在 GB/T 5261 中规定,假定与它们有关的数据是按正向顺序处理的,当它们包含在不按正向顺序处理的数据串中,或包含在供固定记录处理的格式化数据中时,可能会产生不希望有的影响,或者可能需要附加特定的处理,以保证它们产生预期的功能。

2 一致性和实现

2.1 一致性

2.1.1 信息交换的一致性

若编码字符数据表(CC 数据元素)内所有字符的编码表示与本标准 8.1 的要求一致,则交换用的编码信息内的 CC 数据元素具有与本标准的一致性。

一致性要求应按 8.2 和 8.3 所采用的版本标识。

2.1.2 设备的一致性

若设备与 2.1.2.1 以及与 2.1.2.2 和 2.1.2.3 之一或两者一致,则它具有与本标准的一致性。一致性要求应与所采用的版本相一致。

2.1.2.1 设备的描述

与本标准一致的设备应是标识方式描述的对象,当字符对用户有用时,如分别按 2.1.2.2 和 2.1.2.3 中的规定,用户可以向设备供给字符,或者可以识别字符。

2.1.2.2 始发设备

始发设备应允许它的用户提供来自所采用版本的任何字符序列,并且应有能力在 CC 数据元素内传送它的编码表示。

2.1.2.3 接收设备

接收设备应有能力接收和解释任何在 CC 数据元素之内的字符编码表示,并且它与 2.1.1 一致。接收设备应使相应的字符对它的用户有效,以这种方式,用户能从所采用的版本中标识字符,并能将这些

字符相互区分。

2.2 实现

使用本字符集需要规定它在各种媒体上的表示方法,这些媒体可能包括穿孔纸带、穿孔卡片、磁和光的可交换媒体及传输信道,因而允许通过记录在物理媒体上的中间记录间接地进行数据交换,或者通过各种设备(例如输入输出设备和计算机)在本地的连接或者通过数据传输设备进行数据交换。

在物理媒体上和传输时,实现这种编码字符集,考虑其差错检验应由其他标准规定。

3 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/T 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/T 5261—1994 信息处理 七位和八位编码字符集用的控制功能(idt ISO 6429:1988)

GB/T 9387.1—1998 信息技术 开放系统互连 基本参考模型 第1部分:基本模型(idt ISO/IEC 7498-1:1994)

GB/T 11383—1989 信息处理 信息交换用八位代码结构和编码规则(idt ISO/DIS 9281)

GB/T 12054—1989 数据处理 转义序列的登记规程(eqv ISO 2375:1985)

4 定义

本标准采用下列定义。

4.1 现行位置 active position

表示下一个图形字符的图形符号成像的字符位置,或者与下一个要执行的控制功能有关的位置。

注1:一般情况下,现行位置在显示器上用光标指示出来。

4.2 位组 bit combination

用于表示字符的若干个二进制位的有序集合。

4.3 字符 character

用于组织、控制、表示数据的元素集合中的一个元素。

4.4 字符位置 character position

正在成像或能够成像一个图形符号的显示位置。

4.5 编码字符集 coded character set; 代码 code

一组明确的规则,它建立字符集和该字符集内的字符与其位组之间一一对应的关系。

4.6 编码字符数据元素(CC数据元素) coded-character-data-element(CC-data-element)

交换信息的元素,它由一个或多个已标识的编码字符集标准的一些字符的编码表示序列组成。

注2:在通信环境中,开放系统互连的参考模型(GB/T 9387.1),CC数据元素将构成信息的全部或一部分,它符合那个标准中定义的表示协议数据单元(PPDU)。

注3:当信息交换用可交换媒体的方式完成时,CC数据元素构成信息的全部或一部分,它符合用户数据,而它在格式化和初始化期间不记录。

4.7 代码扩充 code extension

对未包括在给定字符集内的字符进行编码的技术。

4.8 代码表 code table

表示代码中每个字符与其指定位组相对应的表。

4.9 控制字符 control character

一种控制功能,它的编码表示由单个位组组成。

4.10 控制功能 control function

影响数据的记录、处理、传输或解释的一种动作或作用,其编码表示由一个或多个位组组成。

4.11 设备 device

信息处理设备的组成部分,它能发送和(或)接收 CC 数据元素内的编码信息。

注 4: 它可以是习惯意义上的输入输出设备,或者像应用程序或网关功能那样的进程。

4.12 转义序列 escape sequence

在代码扩充过程中,为了控制的目的使用若干位组,这些位组中的第一个是表示控制功能的 ESCAPE。

4.13 终止字节 final byte

终止转义序列或控制序列的那个位组。

4.14 图形字符 graphic character

不同于控制功能的一类字符,它有可见的表示,通常是手写的、打印的或显示的,并且它有一个或多个位组组成的编码表示。

4.15 图形符号 graphic symbol

图形字符或控制功能的可见表示。

4.16 用户 user

请求由设备提供服务的人或其他实体。

注 5: 若“设备”是代码转换器或网关功能,则这个实体可以是进程,如:应用程序。

注 6: 用户供给的或可用的字符,可取设备本地代码的形式,或者可取 2.1.2 所规定的那种非习惯的可见表示形式。

5 记法、代码表和名称

5.1 记法

七位代码表位组的各位用 b_7 、 b_6 、 b_5 、 b_4 、 b_3 、 b_2 和 b_1 标识,这里 b_7 是最高位,或最高有效位,而 b_1 是最低位,或最低有效位,见表 1。

在二进制记法中,利用下列权值对各位进行加权,则可以把位组解释成代表 0~127 范围内的整数。

表 1 二进制位的标识和权值

位	b_7	b_6	b_5	b_4	b_3	b_2	b_1
权值	64	32	16	8	4	2	1

本标准中,位组用 x/y 形式的记数法表示。这里 x 是 0~7 范围内的数, y 是 0~15 范围内的数。 x/y 形式的记数法和 $b_7 \sim b_1$ 位所组成的位组之间的相应关系如下:

—— x 是 b_7 、 b_6 和 b_5 表示的数,赋予 b_7 、 b_6 和 b_5 的权值分别为 4、2 和 1。

—— y 是 b_4 、 b_3 、 b_2 和 b_1 表示的数,赋予 b_4 、 b_3 、 b_2 和 b_1 的权值分别为 8、4、2 和 1。

x/y 形式表示的记数法和使用代码表位置是等同的,其中 x 是列号, y 是行号(见 5.2)。

5.2 代码表

七位代码表由安排在 8 列和 16 行中的 128 个位置组成,列号是 0~7,而行号是 0~15。

代码表的位置用 x/y 的形式来标识,其中 x 是列号,而 y 是行号。

代码表的各个位置同代码的各个位组一一对应。

用 x/y 形式表示代码表位置的记法与对应位组的记法是相同的。

5.3 名称

本标准对每个字符指定了一个名称。此外,对每个控制字符、间隔字符和抹掉字符规定一个缩写符号,并对每个图形字符规定一个图形符号。通常用汉字、大写字母书写字符的名称,只有大写字母和数字可用于缩写符号。有意将这些缩写和约定保持在文本的所有解释中。

所选图形字符的名称反映了它们的习惯含义,但是,本标准不定义也不限制这些图形字符的含义,

也不规定图形字符成像时的特殊风格或字型设计。

6 编码字符集规范

6.1 结构

本标准编码字符集应具有下面的结构：

- 32 个控制字符的 C0 控制字符集；
- 图形字符“间隔”(SP)；
- 94 个图形字符的 G0 图形字符集；
- 字符“抹掉”(DEL)。

编码字符集的结构见图 1。

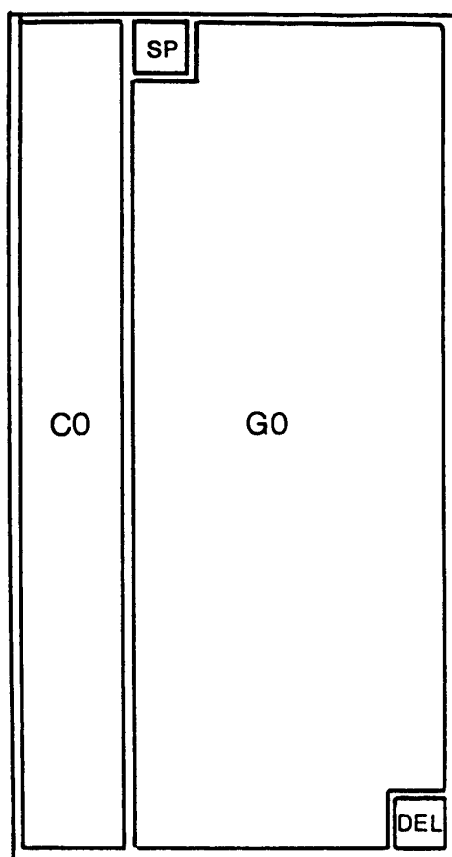


图 1 七位编码字符集的结构

6.2 控制字符

C0 集的控制字符应由附录 A 中所规定的位组表示。

6.3 字符“间隔”

字符“间隔”的缩写是 SP, 并且它由位组 2/0 表示。

这个字符是图形字符, 它是由不存在的图形符号组成的可见表示。

6.4 图形字符

94 个位组 2/1 到 7/14 用于表示图形字符, 按 6.4.1 和 6.4.2 中的规定。

所有的图形字符都是进格字符, 即它们引起现行位置按一个一个的字符位置前进。

6.4.1 唯一图形字符分配

在 2/1 到 2/3, 2/5 到 7/14 的 91 个位置上, 对每个位置分配一个唯一的图形字符。这些图形字符在表 2 中规定。

表 2 唯一图形字符的分配

图 形 符 号	名 称	编 码 表 示
!	感叹号	2/1
"	双引号	2/2
#	数码记号	2/3
%	百分比	2/5
&	和	2/6
'	撇号	2/7
(左圆括号	2/8
)	右圆括号	2/9
*	星号	2/10
+	正号	2/11
,	逗号	2/12
-	负号	2/13
.	句点	2/14
/	斜线	2/15
0	数字〇	3/0
1	数字一	3/1
2	数字二	3/2
3	数字三	3/3
4	数字四	3/4
5	数字五	3/5
6	数字六	3/6
7	数字七	3/7
8	数字八	3/8
9	数字九	3/9
:	冒号	3/10
;	分号	3/11
<	小于记号	3/12
=	等于记号	3/13
>	大于记号	3/14
?	问号	3/15
@	商用(单价)记号	4/0
A	拉丁大写字母 A	4/1
B	拉丁大写字母 B	4/2
C	拉丁大写字母 C	4/3
D	拉丁大写字母 D	4/4
E	拉丁大写字母 E	4/5
F	拉丁大写字母 F	4/6
G	拉丁大写字母 G	4/7
H	拉丁大写字母 H	4/8
I	拉丁大写字母 I	4/9
J	拉丁大写字母 J	4/10

表 2(续)

图 形 符 号	名 称	编 码 表 示
K	拉丁大写字母 K	4/11
L	拉丁大写字母 L	4/12
M	拉丁大写字母 M	4/13
N	拉丁大写字母 N	4/14
O	拉丁大写字母 O	4/15
P	拉丁大写字母 P	5/0
Q	拉丁大写字母 Q	5/1
R	拉丁大写字母 R	5/2
S	拉丁大写字母 S	5/3
T	拉丁大写字母 T	5/4
U	拉丁大写字母 U	5/5
V	拉丁大写字母 V	5/6
W	拉丁大写字母 W	5/7
X	拉丁大写字母 X	5/8
Y	拉丁大写字母 Y	5/9
Z	拉丁大写字母 Z	5/10
[左方括号	5/11
\	反斜线	5/12
]	右方括号	5/13
^	向上箭头	5/14
—	下横线	5/15
`	右撇号	6/0
a	拉丁小写字母 A	6/1
b	拉丁小写字母 B	6/2
c	拉丁小写字母 C	6/3
d	拉丁小写字母 D	6/4
e	拉丁小写字母 E	6/5
f	拉丁小写字母 F	6/6
g	拉丁小写字母 G	6/7
h	拉丁小写字母 H	6/8
i	拉丁小写字母 I	6/9
j	拉丁小写字母 J	6/10
k	拉丁小写字母 K	6/11
l	拉丁小写字母 L	6/12
m	拉丁小写字母 M	6/13
n	拉丁小写字母 N	6/14
o	拉丁小写字母 O	6/15
p	拉丁小写字母 P	7/0
q	拉丁小写字母 Q	7/1
r	拉丁小写字母 R	7/2
s	拉丁小写字母 S	7/3
t	拉丁小写字母 T	7/4
u	拉丁小写字母 U	7/5
v	拉丁小写字母 V	7/6

表 2(完)

图 形 符 号	名 称	编 码 表 示
w	拉丁小写字母 W	7/7
x	拉丁小写字母 X	7/8
y	拉丁小写字母 Y	7/9
z	拉丁小写字母 Z	7/10
{	左花括号	7/11
	竖线	7/12
}	右花括号	7/13
~	上波浪号	7/14

6.4.2 替换的图形字符分配

两个替换的图形字符分配到 2/4,分别用于中国的国家版本和国际参考版本。这两个字符在表 3 中规定。

表 3 替换图形字符的分配

图形符号	名 称	编码表示	版 本
¥	元记号	2/4	中国国家版本
\$	元记号	2/4	国际参考版本(IRV)

“元记号”字符“¥”或“元记号”字符“\$”应分配到位组 2/4(见第 8 章)。

在国际交往中,除发送者和接收者双方有协定外,图形符号¥和\$不指定作为特定国家的货币。

国内的交往中,在通常情况下,图形符号“¥”可作为人民币货币符的元记号。若有特殊的规定除外。

6.5 字符“抹掉”

字符“抹掉”的缩写是 DEL,它由位组 7/15 表示。DEL 原来是用于清除或擦去穿孔带上错误的无用的字符。DEL 可以被作为媒体填充或时间占空使用。DEL 字符可以插入到或移出数据流,而不影响那个流的信息控制,但这样的作用可能影响信息的布局和(或)设备的控制。

7 组合图形字符

虽然本标准中规定的所有图形字符是进格字符,但是,使用“退格”或“回车”字符在同一字符位置上,两个或多个图形字符成像是可能的。

例如“斜线”和“等于记号”可组合成“不等”的图像。字符“下横线”既可以作为一个独立的字符使用,也可以与其他字符结合表示“有下横线”的图形显现式样。

可以利用发音符组合成带音调的字母,用一个三字符序列进行这种组合。其第一个或最后一个字符是要标音调的字母,第二个字符是“退格”,而且“双引号”、“撇号”或“逗号”可利用“退格”和一个字母连用,分别形成一个带有发音的字母、带有高音符的字母或下带变音符“,”的字母。

8 编码字符集的版本

8.1 概述

为了使用信息交换的七位编码字符集,运用第 6 章中的选项是必需的。

- 依据附录 A 规定 C0 控制字符集;
- 规定 G0 集,2/4 位组的图形符号按 6.4.2 规定。

8.2 中国版本

依据国际标准 ISO/IEC 646 的基本代码表定义国家版本是各国家标准团体的职责,按照中国实际需要,2/4 位置的图形符号规定为“¥”,即构成了中国版本(见表 4)。

8.3 国际参考版本(IRV)

不要求使用国家版本时,这个版本才是有效的。在信息交换中,IRV 是在数据的发送者和接收者双方有协定的情况下使用(见第 10 章)。分配给 IRV 的 G0 集的图形字符应是表 5 中所规定的。若使用组合字符,也需要数据收发双方有协定。

若使用 IRV 的 C0 集,也应按照附录 A 中的第 A2 章的规定。

9 版本的标识

9.1 标识的目的和上下文

同本标准版本一致的 CC 数据元素,应由发送者和接收者之间交换的编码信息之全部或组合单元的一部分构成。发送者采用本标准版本的标识对接收者也应有效的。将这种标识通知到接收者的路由不在本标准中规定。

但是,用 CC 数据元素标识的编码表示构成被交换信息的一部分,这对于编码信息交换用的某些标准是允许的,或者是要求的。本章为本标准版本的标识规定了编码表示。这种编码表示形成了标识数据元素的全部或一部分,它可以包含在信息交换中,与相关的标准一致。

9.2 版本的标识

本标准版本的标识应由标识集组成,有一个 C0 集和 G0 集构成该版本。在该集合中的每个标识由分配的典型转义序列示出如下:

ESC 2/1 F 标识 C0 集。

ESC 2/8 F 标识 G0 集。

这些转义序列中最后的终止字节 F 应从 GB/T 12054 指明的登记处获得。若 C0 集是空的,则标识应是相同的转义序列,它最后的字节 F 是 7/14。

IRV 应由下面的转义序列标识:

若提供 C0 集,则用登记 ISO IR No. 1;ESC 2/1 4/0。

若不提供 C0 集,则用 ESC 2/1 7/14。

对于 G0 集,用登记 ISO IR No. 6;ESC 2/8 4/2。

中国版本应用下面的转义序列标识:

若提供 C0 集,则用登记的 ISO IR No. 1;ESC 2/1 4/0。

对于 G0 集,用登记的转义序列;ESC 2/8 5/4。

10 表 4 和表 5 的说明

表 5 的 2/4 位置是图形符号“\$”,但它不代表特定国家的货币符号,该版本是国际参考版本(IRV)。

位组 0/0 到 1/15 是为控制字符保留的,这些位置是空的。位组 2/0 到 7/15 表示图形符号,或者缩写。

表 4 是一个包含中国版本 G0 集而没有 C0 集的代码表。

表 5 是一个包含 IRV 的 G0 集而没有 C0 集的代码表。

表 4 中国的 G0 版本

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

表5 IRV 的 G0 版本

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

附 录 A
(标准的附录)
C0 集规范

A1 GB/T 1988 版本的 C0 集

GB/T 1988 版本的 C0 集应是 GB/T 5261 中的 C0 集。这些控制功能应按它们在 GB/T 5261 中的定义使用。

A2 IRV 的 C0 集

若用国际参考版本(IRV)的 C0 集,则它应是 GB/T 5261 的 C0 集。

A3 表 A1

表 A1 示出了 GB/T 5261 的 C0 集。这些控制功能的定义见 GB/T 5261。

表 A1 GB 5261 的 C0 集

					b7	0	0
					b6	0	0
					b5	0	1
						0	1
b4	b3	b2	b1				
0	0	0	0	0	NUL	DLE	
0	0	0	1	1	SOH	DC1	
0	0	1	0	2	STX	DC2	
0	0	1	1	3	ETX	DC3	
0	1	0	0	4	EOT	DC4	
0	1	0	1	5	ENQ	NAK	
0	1	1	0	6	ACK	SYN	
0	1	1	1	7	BEL	ETB	
1	0	0	0	8	BS	CAN	
1	0	0	1	9	HT	EM	
1	0	1	0	10	LF	SUB	
1	0	1	1	11	VT	ESC	
1	1	0	0	12	FF	IS.	
1	1	0	1	13	CR	IS.	
1	1	1	0	14	SO	IS.	
1	1	1	1	15	SI	IS.	

附 录 B

(提示的附录)

本标准第二版(1989)与现行(第三)版之间的差别

- B1** 所有代码标准采用的一致性条款都引入了新文本。
 - B2** 所有定义的用词都与 GB/T 11383 和 ISO/IEC 10538 相同。
 - B3** 在 IRV 的 G0 集内,字符“元记号”代替 2/4 位置上的字符“通用货币记号”。
 - B4** 这个 C0 集在标准的附录 A 中示出。32 个控制字符的定义引用 GB/T 5261。
表 4 和表 5 没有在 0 列和 1 列示出任何特定的控制字符。
-