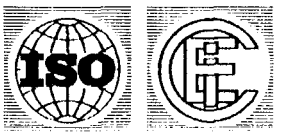
INTERNATIONAL STANDARD

ISO/IEC 4873

Third edition
1991-12-15

Information technology — ISO 8-bit code for information interchange — Structure and rules for implementation

Technologies de l'information — Code ISO à 8 éléments pour l'échange d'informations — Structure et règles de matérialisation



Reference number
ISO/IEC 4873:1991(E)

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS

Not for Resale

Contents

	Page
1 Scope	1
2 Conformance and implementation	1
2.1 Conformance	1
2.1.1 Conformance of information interchange	1
2.1.2 Conformance of devices	1
2.2 Implementation	2
3 Normative references	2
4 Definitions	2
4.1 active position	2
4.2 bit combination	2
4.3 byte 2	
4.4 character	2
4.5 character position	2
4.6 coded-character-data-element (CC-data-element)	2
4.7 coded character set	3
4.8 code extension	3
4.9 code table	3
4.10 control character	3
4.11 control function	3
4.12 device	3
4.13 escape sequence	3
4.14 Final Byte	3
4.15 graphic character	3
4.16 graphic symbol	3
4.17 repertoire	3
4.18 user 3	
5 Notation, code table and names	3
5.1 Notation	3
5.2 Code table	4
5.3 Names	4
6 Structure of the 8-bit code	4
6.1 Elements of the 8-bit code	4
6.2 Identification of the elements of the 8-bit code	5

© ISO/IEC 1991

All rights reserved. No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the publisher.

ISO/IEC Copyright Office • Case postale 56 • CH-1211 Genève 20 • Switzerland

Printed in Switzerland

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS

Not for Resale

6.3 Invocation 5

6.3.1 C0 set 5
6.3.2 Character SPACE 5
6.3.3 G0 set 5
6.3.4 Character DELETE 5
6.3.5 C1 set 5
6.3.6 G1 set 5
6.3.7 G2 set 5
6.3.8 G3 set 5

7 Specification of the characters of the 8-bit code 5

7.1 C0 set 6
7.2 Character ESCAPE 6
7.3 Character SPACE 6
7.4 G0 set 6
7.5 Character DELETE 9
7.6 C1 set 9
7.7 G1 set 9
7.8 G2 set 9
7.9 G3 set 10
7.10 Summary of the specification of the 8-bit code 10

8 Levels 10

8.1 Level 1 10
8.2 Level 2 10
8.3 Level 3 11

9 Version of the 8-bit code 11

9.1 Contents of a version 11
9.2 Unique coding of characters 11

10 Identification of version and level 11

10.1 Purpose and context of identification 11
10.2 Identification of level 12
10.3 Identification of a version 12
10.4 Switching from one version to another 12
10.5 Switching from one level to another 12

Annexes

A Restrictions applicable to the C0 and C1 sets 16

B Shift functions 17

C Composite graphic characters 18

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

D Use of bit combinations 00/14 and 00/15 19

E Main differences between the 2nd edition (1986) and the present (third) edition of this International Standard 19

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work

In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

International Standard ISO/IEC 4873 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*.

This third edition cancels and replaces the second edition (ISO 4873 : 1986), which has been technically revised.

Annex A forms an integral part of this International Standard. Annexes B, C, D and E are for information only.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

Information technology — ISO 8-bit code for information interchange — Structure and rules for implementation

1 Scope

This International Standard specifies an 8-bit code derived from, and compatible with, the 7-bit coded character set specified in ISO/IEC 646.

The characteristics of this code are also in conformance with the code extension techniques specified in ISO 2022.

This International Standard specifies an 8-bit code with a number of options. It also provides guidance on how to exercise the options to define specific versions.

This code is primarily intended for general information interchange within an 8-bit environment among data processing systems and associated equipment, and within data communication systems. The need for graphic characters and control functions in data processing has also been taken into account.

The code includes the 10 digits as well as the 52 small and capital letters of the basic Latin alphabet and may include accented letters, special Latin letters and/or the letters of one or several non-Latin alphabet(s).

2 Conformance and implementation

2.1 Conformance

2.1.1 Conformance of information interchange

A coded-character-data-element (CC-data-element) within coded information for interchange is in conformance with a version of this International Standard if all the coded representations of characters within that CC-data-element conform to the requirements of clause 9.

A claim of conformance shall identify the version adopted.

2.1.2 Conformance of devices

A device is in conformance with this International Standard if it conforms to the requirements of 2.1.2.1, and either or both of 2.1.2.2 and 2.1.2.3. A claim of conformance shall identify the document which contains the description specified in 2.1.2.1, and shall identify the version adopted.

2.1.2.1 Device description

A device that conforms to this International Standard shall be the subject of a description that identifies the means by which the user may supply characters to the device, or may recognize them when they are made available to him, as specified respectively in 2.1.2.2 and 2.1.2.3.

2.1.2.2 Originating devices

An originating device shall allow its user to supply any sequence of characters from the version adopted, and shall be capable of transmitting their coded representations within a CC-data-element.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

2.1.2.3 Receiving devices

A receiving device shall be capable of receiving and interpreting any coded representations of characters that are within a CC-data-element, and that conform to 2.1.1, and shall make the corresponding characters available to its user in such a way that the user can identify them from among those of the version adopted, and can distinguish them from each other.

2.2 Implementation

The use of this code requires definitions of its implementation in various media. For example, these could include punched tapes, punched cards, magnetic and optical media and transmission channels, thus permitting interchange of data to take place either indirectly by means of an intermediate recording in a physical medium, or by local connection of various units (such as input and output devices and computers) or by means of data transmission equipment.

The implementation of this code in physical media and for transmission, taking into account the need for error checking, is the subject of other International Standards.

3 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this International Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the standards listed below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO/IEC 646: 1991 *Information technology - ISO 7-bit coded character set for information interchange.*

ISO 2022: 1986 *Information processing - ISO 7-bit and 8-bit coded character sets - Code extension techniques.*

ISO/IEC 6429:[1] *Information technology - Control functions for 7-bit and 8-bit coded character sets.*

ISO/IEC 10367: 1991 *Information technology - Standardized coded graphic character sets for use in 8-bit codes.*

ISO International Register of Coded Character Sets to be Used with Escape Sequences (ISO 2375).

4 Definitions

For the purpose of this International Standard the following definitions apply.

4.1 active position : The character position which is to image the graphic symbol representing the next graphic character or relative to which the next control function is to be executed.

NOTE - In general, the active position is indicated in a display by a cursor.

4.2 bit combination : An ordered set of bits used for the representation of characters.

4.3 byte : A bit string that is operated upon as a unit.

4.4 character : A member of a set of elements used for the organization, control or representation of data.

4.5 character position : The portion of a display that is imaging or is capable of imaging a graphic symbol.

4.6 coded-character-data-element (CC-data-element) : An element of interchanged information that is specified to consist of a sequence of coded representations of characters, in accordance with one or more identified standards for coded character sets.

---

1) To be published.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

NOTES

1. In a communication environment according to the Reference Model for Open Systems Interconnection (ISO 7498), a CC-data-element will form all or part of the information that corresponds to the Presentation-Protocol-Data-Unit (PPDU) defined in that International Standard.

2. When information interchange is accomplished by means of interchangeable media, a CC-data-element will form all or part of the information that corresponds to the user data, and not that recorded during formatting and initialization.

4.7 coded character set; code : A set of unambiguous rules that establishes a character set and the one-to-one relationship between the characters of the set and their bit combinations.

4.8 code extension : The techniques for the encoding of characters that are not included in the character set of a given code.

4.9 code table : A table showing the character allocated to each bit combination in a code.

4.10 control character : A control function the coded representation of which consists of a single bit combination.

4.11 control function : An action that affects the recording, processing, transmission, or interpretation of data, and that has a coded representation consisting of one or more bit combinations.

4.12 device : A component of information processing equipment which can transmit, and/or receive, coded information within CC-data-elements.

NOTE - It may be an input/output device in the conventional sense, or a process such as an application program or gateway function.

4.13 escape sequence : A string of bit combinations that is used for control purposes in code extension procedures. The first of these bit combinations represents the control function ESCAPE.

4.14 Final Byte : The bit combination that terminates an escape sequence or a control sequence.

4.15 graphic character : A character, other than a control function, that has a visual representation normally handwritten, printed or displayed, and that has a coded representation consisting of one or more bit combinations.

4.16 graphic symbol : A visual representation of a graphic character or of a control function.

4.17 repertoire : A specified set of characters that are represented by means of one or more bit combinations of a coded character set.

4.18 user : A person or other entity that invokes the services provided by a device.

NOTES

1. This entity may be a process such as an application program if the "device" is a code convertor or a gateway function, for example.

2. The characters, as supplied by the user or made available to him, may be in the form of codes local to the device, or of non-conventional visible representations, provided that 2.1.2 above is satisfied.

5 Notation, code table and names

5.1 Notation

The bits of the bit combinations of the 8-bit code are identified by b_8, b_7, b_6, b_5, b_4, b_3, b_2 and b_1, where b_8 is the highest-order, or most-significant bit, and b_1 is the lowest-order, or least-significant, bit.

The bit combinations may be interpreted to represent integers in the range 0 to 255 in binary notation by attributing the following weights to the individual bits:

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

Bit	b_8	b_7	b_6	b_5	b_4	b_3	b_2	b_1
Weight	128	64	32	16	8	4	2	1

In this International Standard, the bit combinations are identified by notations of the form xx/yy, where xx and yy are numbers in the range 00 to 15. The correspondence between the notations of the form xx/yy and the bit combinations consisting of the bits b_8 to b_1, is as follows:

- xx is the number represented by b_8, b_7, b_6 and b_5 where these bits are given the weights 8, 4, 2 and 1 respectively;
- yy is the number represented by b_4, b_3, b_2 and b_1 where these bits are given the weights 8, 4, 2 and 1 respectively.

The notations of the form xx/yy are the same as the ones used to identify code table positions, where xx is the column number and yy is the row number (see 5.2).

5.2 Code table

An 8-bit code table consists of 256 positions arranged in 16 columns and 16 rows. The columns and rows are numbered 00 to 15.

The code table positions are identified by notations of the form xx/yy, where xx is the column number and yy is the row number.

The positions of the code table are in one-to-one correspondence with the bit combinations of the code. The notation of a code table position, of the form xx/yy, is the same as that of the corresponding bit combination.

5.3 Names

This International Standard assigns one name to each character. In addition, it specifies an acronym for control characters and for the characters SPACE and DELETE, and a graphic symbol for each graphic character. By convention, only capital letters, and hyphen are used for writing the names of the characters. For acronyms only capital letters, space and digit are used. It is intended that the acronyms and this convention be retained in all translations of the text.

The names chosen to denote graphic characters are intended to reflect their customary meaning. However, this International Standard does not define and does not restrict the meanings of graphic characters. Neither does it specify a particular style or font design for the graphic characters when imaged.

6 Structure of the 8-bit code

6.1 Elements of the 8-bit code

The 8-bit code consists of the following parts (see figure 1).

a) **A C0 set**

A set of up to 30 control characters represented by bit combinations 00/00 to 01/15, except 00/14 and 00/15 which shall be unused.

b) **The character SPACE**

A graphic character represented by bit combination 02/00.

c) **A G0 set**

A set of 94 graphic characters represented by bit combinations 02/01 to 07/14.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

d) **The character DELETE**

A character represented by bit combination 07/15.

e) **A C1 set**

A set of up to 32 control characters represented by bit combinations 08/00 to 09/15.

f) **A G1 set**

A set of up to 96 graphic characters represented by bit combinations 10/00 to 15/15.

g) **A G2 set**

A set of up to 96 graphic characters.

h) **A G3 set**

A set of up to 96 graphic characters.

6.2 Identification of the elements of the 8-bit code

The method of identification of the code elements listed in 6.1 is specified in clause 10.

6.3 Invocation

6.3.1 C0 set

The identification of the C0 set also invokes that set.

6.3.2 Character SPACE

The character SPACE shall be represented by bit combination 02/00. It is not explicitly invoked.

6.3.3 G0 set

The G0 set shall be as specified in 7.4. It is not explicitly invoked.

6.3.4 Character DELETE

The character DELETE shall be represented by bit combination 07/15. It is not explicitly invoked.

6.3.5 C1 set

The identification of the C1 set also invokes that set.

6.3.6 G1 set

The identification of the G1 set also invokes that set. The locking-shift function LS1R shall also invoke the G1 set.

6.3.7 G2 set

Either the set as a whole shall be invoked by the locking-shift function LS2R (see annex B) into columns 10 to 15, or individual characters of it shall be invoked by means of the single-shift function SS2, (see 7.6).

6.3.8 G3 set

Either the set as a whole shall be invoked by the locking-shift function LS3R (see annex B) into columns 10 to 15, or individual characters of the set shall be invoked by means of the single-shift function SS3 (see 7.6).

7 Specification of the characters of the 8-bit code

The use of control functions such as BACKSPACE or CARRIAGE RETURN, for the coded representation of composite characters is prohibited by this International Standard (see annex C).

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

7.1 C0 set

The requirements for the C0 set are

- bit combinations 00/14 and 00/15 shall not be used (see annex D);
- the control character ESCAPE shall be represented by bit combination 01/11;
- any control characters can be allocated to the other bit combinations subject to the restrictions specified in annex A.

NOTE - A C0 set comprising only ESCAPE represented by bit combination 01/11 has been registered (Registration ISO-IR No. 104), and is identified by ESC 02/01 04/07.

7.2 Character ESCAPE

ESCAPE is a control character used to form escape sequences. In this International Standard the use of escape sequences is specified in clause 10.

Table 1 - ESCAPE

Acronym	Name	Coded representation
ESC	ESCAPE	01/11

7.3 Character SPACE

A graphic character having a visual representation consisting of the absence of a graphic symbol. It causes the active position to be advanced by one character position.

Table 2 - SPACE

Acronym	Name	Coded representation
SP	SPACE	02/00

7.4 G0 set

The 94 bit combinations 02/01 to 07/14 are used to represent graphic characters. All graphic characters allocated to bit combinations in the range 02/01 to 07/14 are spacing characters, that is they cause the active position to advance by one character position. The graphic characters allocated by this International Standard to these 94 bit combinations are specified in table 3.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

Table 3 - Graphic characters of the G0 set

Graphic Symbol	Name	Coded representation
!	EXCLAMATION MARK	02/01
"	QUOTATION MARK	02/02
#	NUMBER SIGN	02/03
$	DOLLAR SIGN	02/04
%	PERCENT SIGN	02/05
&	AMPERSAND	02/06
'	APOSTROPHE	02/07
(	LEFT PARENTHESIS	02/08
)	RIGHT PARENTHESIS	02/09
*	ASTERISK	02/10
+	PLUS SIGN	02/11
,	COMMA	02/12
-	HYPHEN-MINUS	02/13
.	FULL STOP	02/14
/	SOLIDUS	02/15
0	DIGIT ZERO	03/00
1	DIGIT ONE	03/01
2	DIGIT TWO	03/02
3	DIGIT THREE	03/03
4	DIGIT FOUR	03/04
5	DIGIT FIVE	03/05
6	DIGIT SIX	03/06
7	DIGIT SEVEN	03/07
8	DIGIT EIGHT	03/08
9	DIGIT NINE	03/09
:	COLON	03/10
;	SEMICOLON	03/11
<	LESS-THAN SIGN	03/12
=	EQUALS SIGN	03/13
>	GREATER-THAN SIGN	03/14
?	QUESTION MARK	03/15
	COMMERCIAL AT	04/00
A	LATIN CAPITAL LETTER A	04/01
B	LATIN CAPITAL LETTER B	04/02
C	LATIN CAPITAL LETTER C	04/03
D	LATIN CAPITAL LETTER D	04/04
E	LATIN CAPITAL LETTER E	04/05
F	LATIN CAPITAL LETTER F	04/06
G	LATIN CAPITAL LETTER G	04/07
H	LATIN CAPITAL LETTER H	04/08
I	LATIN CAPITAL LETTER I	04/09
J	LATIN CAPITAL LETTER J	04/10
K	LATIN CAPITAL LETTER K	04/11
L	LATIN CAPITAL LETTER L	04/12
M	LATIN CAPITAL LETTER M	04/13
N	LATIN CAPITAL LETTER N	04/14
O	LATIN CAPITAL LETTER O	04/15
P	LATIN CAPITAL LETTER P	05/00

(continued)

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

Table 3 - Graphic characters of the G0 set

Graphic Symbol	Name	Coded representation
Q	LATIN CAPITAL LETTER Q	05/01
R	LATIN CAPITAL LETTER R	05/02
S	LATIN CAPITAL LETTER S	05/03
T	LATIN CAPITAL LETTER T	05/04
U	LATIN CAPITAL LETTER U	05/05
V	LATIN CAPITAL LETTER V	05/06
W	LATIN CAPITAL LETTER W	05/07
X	LATIN CAPITAL LETTER X	05/08
Y	LATIN CAPITAL LETTER Y	05/09
Z	LATIN CAPITAL LETTER Z	05/10
[	LEFT SQUARE BRACKET	05/11
\	REVERSE SOLIDUS	05/12
]	RIGHT SQUARE BRACKET	05/13
^	CIRCUMFLEX ACCENT	05/14
_	LOW LINE	05/15
`	GRAVE ACCENT	06/00
a	LATIN SMALL LETTER A	06/01
b	LATIN SMALL LETTER B	06/02
c	LATIN SMALL LETTER C	06/03
d	LATIN SMALL LETTER D	06/04
e	LATIN SMALL LETTER E	06/05
f	LATIN SMALL LETTER F	06/06
g	LATIN SMALL LETTER G	06/07
h	LATIN SMALL LETTER H	06/08
i	LATIN SMALL LETTER I	06/09
j	LATIN SMALL LETTER J	06/10
k	LATIN SMALL LETTER K	06/11
l	LATIN SMALL LETTER L	06/12
m	LATIN SMALL LETTER M	06/13
n	LATIN SMALL LETTER N	06/14
o	LATIN SMALL LETTER O	06/15
p	LATIN SMALL LETTER P	07/00
q	LATIN SMALL LETTER Q	07/01
r	LATIN SMALL LETTER R	07/02
s	LATIN SMALL LETTER S	07/03
t	LATIN SMALL LETTER T	07/04
u	LATIN SMALL LETTER U	07/05
v	LATIN SMALL LETTER V	07/06
w	LATIN SMALL LETTER W	07/07
x	LATIN SMALL LETTER X	07/08
y	LATIN SMALL LETTER Y	07/09
z	LATIN SMALL LETTER Z	07/10
{	LEFT CURLY BRACKET	07/11
|	VERTICAL LINE	07/12
}	RIGHT CURLY BRACKET	07/13
~	TILDE	07/14

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

7.5 Character DELETE

DEL was originally used to erase or obliterate an erroneous or unwanted character in punched tape. DEL may be used for media-fill or time-fill. DEL characters may be inserted into, or removed from, a data stream without affecting the information content of that stream, but such action may affect the information layout and/or the control of equipment.

Table 4 - DELETE

Acronym	Name	Coded representation
DEL	DELETE	07/15

7.6 C1 set

The C1 set is available for up to 32 control characters in addition to those provided by the C0 set. It shall not include any of the control characters of the C0 set of ISO 6429.

No specific control characters are allocated to bit combinations 08/00 to 08/13 and 09/00 to 09/15 by this International Standard.

When the single-shift functions SS2 and SS3 are used, they shall be allocated to bit combinations 08/14 and 08/15, respectively, otherwise these bit combinations shall not be used.

NOTE - A C1 set comprising only SS2 and SS3 allocated to these bit combinations has been registered (Registration ISO-IR No. 105), and is identified by ESC 02/02 04/07.

7.7 G1 set

The G1 set shall be either a 94-character or a 96-character set of graphic characters. This set is available for graphic characters in addition to those provided by the G0 set.

Either a unique graphic character shall be allocated to each bit combination or the bit combination shall be declared unused.

The characters of the G1 set are represented by bit combinations 10/01 to 15/14 if the G1 is a 94-character set, or by bit combinations 10/00 to 15/15 if the G1 set is a 96-character set.

7.8 G2 set

The G2 set shall be either a 94-character or a 96-character set of graphic characters. This set is available for graphic characters in addition to those provided by the G0 and the G1 sets.

Either a unique graphic character shall be allocated to each bit combination or the bit combination shall be declared unused.

If the G2 set is a 94-character set, then no characters shall be allocated to bit combinations 10/00 and 15/15.

The characters of the G2 set shall be invoked either by the single-shift function SS2 or by the locking-shift function LS2R.

- When invoked by SS2, each character is represented by the bit combination of SS2 followed by one of the bit combinations in the range 02/01 to 07/14 if the G2 set is a 94-character set, or 02/00 to 07/15 if the G2 set is a 96-character set.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

- When invoked by LS2R, the characters of the G2 set are represented by bit combinations 10/01 to 15/14 if the G2 set is a 94-character set, or by bit combinations 10/00 to 15/15 if the G2 set is a 96-character set.

7.9 G3 set

The G3 set shall be a 94-character or a 96-character set of graphic characters. This set is available for graphic characters in addition to those provided by the G0, the G1 and the G2 sets.

Either a unique graphic character shall be allocated to each bit combination or the bit combination shall be declared unused.

If the G3 set is a 94-character set, then no character shall be allocated to bit combinations 10/00 and 15/15.

The characters of the G3 set shall be invoked either by the single-shift function SS3 or by the locking-shift function LS3R.

- When invoked by SS3, each character is represented by the bit combination of SS3 followed by one of the bit combinations in the range 02/01 to 07/14 if the G3 set is a 94-character set, or 02/00 to 07/15 if the G3 set is a 96-character set.
- When invoked by LS3R, the characters of the G3 set are represented by bit combinations 10/01 to 15/14 if the G3 set is a 94-character set, or by bit combinations 10/00 to 15/15 if the G3 set is a 96-character set.

7.10 Summary of the specification of the 8-bit code

Figure 1 summarizes the specification of the elements of the 8-bit code.

8 Levels

This International Standard specifies three nested levels of implementation.

8.1 Level 1

Level 1, see figure 2, comprises the following facilities:

- a C0 set;
- the character SPACE represented by bit combination 02/00;
- the G0 set;
- the character DELETE represented by bit combination 07/15;
- a C1 set;
- a G1 set.

At Level 1 no shift functions shall be used and the G0 and G1 sets are assumed to be invoked permanently in columns 02 to 07 and 10 to 15, respectively.

At Level 1 the C1 set and/or the G1 set may be empty if there is no requirement for control characters in addition to those provided by the C0 set and/or for graphic characters in addition to those provided by the G0 set.

At Level 1 a version shall not include a G2 or a G3 set.

8.2 Level 2

Level 2, see figure 3, comprises the facilities of Level 1 and in addition to them

- a G2 set the characters of which shall be invoked individually by SS2;
- a G3 set the characters of which shall be invoked individually by SS3.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

At Level 2 no other shift functions shall be used.

The G1 set shall not be empty; either the G2 or the G3 set may be empty but not both.

The C1 set shall not be empty, it shall contain at least the single-shift functions SS2 and SS3.

8.3 Level 3

Level 3, see figure 4, comprises all the facilities of Level 2, with the addition of the following three shift functions:

- LS1R
- LS2R
- LS3R

The G1 set shall not be empty; either the G2 or the G3 set may be, but not both.

The G1, G2 and G3 sets can be invoked explicitly by LS1R, LS2R and LS3R, respectively. Individual characters of the G2 and G3 sets can be invoked by SS2 and SS3, respectively.

The C1 set shall not be empty, it shall contain at least the single-shift functions SS2 and SS3.

9 Version of the 8-bit code

9.1 Contents of a version

A version of the 8-bit code is a coded character set in accordance with clauses 6, 7 and 8, in which a C0 set, a C1 and a G1 set and, optionally, a G2 and a G3 set have been uniquely identified. The level of implementation (clause 8) shall also be identified.

Such a version will generally be the subject of a specification document which states how the above options have been exercised. Such a specification is said to be in accordance with this International Standard.

9.2 Unique coding of characters

In a version the same character may occur in more than one of the G0, G1, G2 and G3 sets. Such a character shall be regarded as the same character as a character in another of those sets if both characters have the same name within the specifications, or ISO International Register entries, that respectively define the two sets.

If the same character has been allocated to more than one of the G0, G1, G2 and G3 sets, either within the set itself or within the character repertoire associated with that set, then that character shall be represented by the coded representation taken from the lowest numbered set (in the sequence G0, G1, G2, G3) in which the character has been allocated.

A coded representation for such a character within one of the other, higher numbered sets shall not be used, even if the higher numbered set is already invoked and the lowest numbered set in which the character is allocated is not currently invoked.

10 Identification of version and level

10.1 Purpose and context of identification

CC-data-elements conforming to a version of this International Standard are intended to form all or part of a composite unit of coded information that is interchanged between a sender and a recipient. The identification of the version of this International Standard that has been adopted by the originator shall also be available to the recipient. The route by which such identification is communicated to the recipient is outside the scope of this International Standard.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

However, some standards for interchange of coded information may permit, or require, that the coded representation of the identification applicable to the CC-data-elements forms a part of the interchanged information. This clause specifies a coded representation for the identification of a version and a level of this International Standard. Such coded representations form all or part of an identifying data element, which may be included in information interchange in accordance with the relevant standard.

10.2 Identification of level

A level of this International Standard shall be identified by means of an announcer sequence from the following list.

ESC 02/00 04/12 shall identify Level 1.
ESC 02/00 04/13 shall identify Level 2.
ESC 02/00 04/14 shall identify Level 3.

10.3 Identification of a version

The identification of a version of this International Standard shall comprise a set of identifications, one for each of the C sets and G sets that constitute the version. Each identification in the set shall consist of a designating escape sequence of the type shown below.

ESC 02/01 F shall identify the C0 set.
ESC 02/02 F shall identify the C1 set.
ESC 02/08 04/02 shall identify the G0 set (Registration ISO-IR No. 6).
ESC 02/09 F or ESC 02/13 F shall identify the G1 set.
ESC 02/10 F or ESC 02/14 F shall identify the G2 set.
ESC 02/11 F or ESC 02/15 F shall identify the G3 set.

The final byte F of these sequences shall be obtained from the International Register (ISO 2375). If any of the C sets or G sets is empty, the identification shall be the same escape sequence in which the final byte F is 07/14.

For a version of this International Standard that conforms to Level 1, no identifications for the G2 and G3 sets shall be included.

NOTE - The designating escape sequence for the G0 set is indicated with the final byte which corresponds to the G0 set of this International Standard. Whilst this designation is not necessary according to this edition of ISO/IEC 4873, it has been kept for reasons of compatibility with the 2nd edition of this International Standard

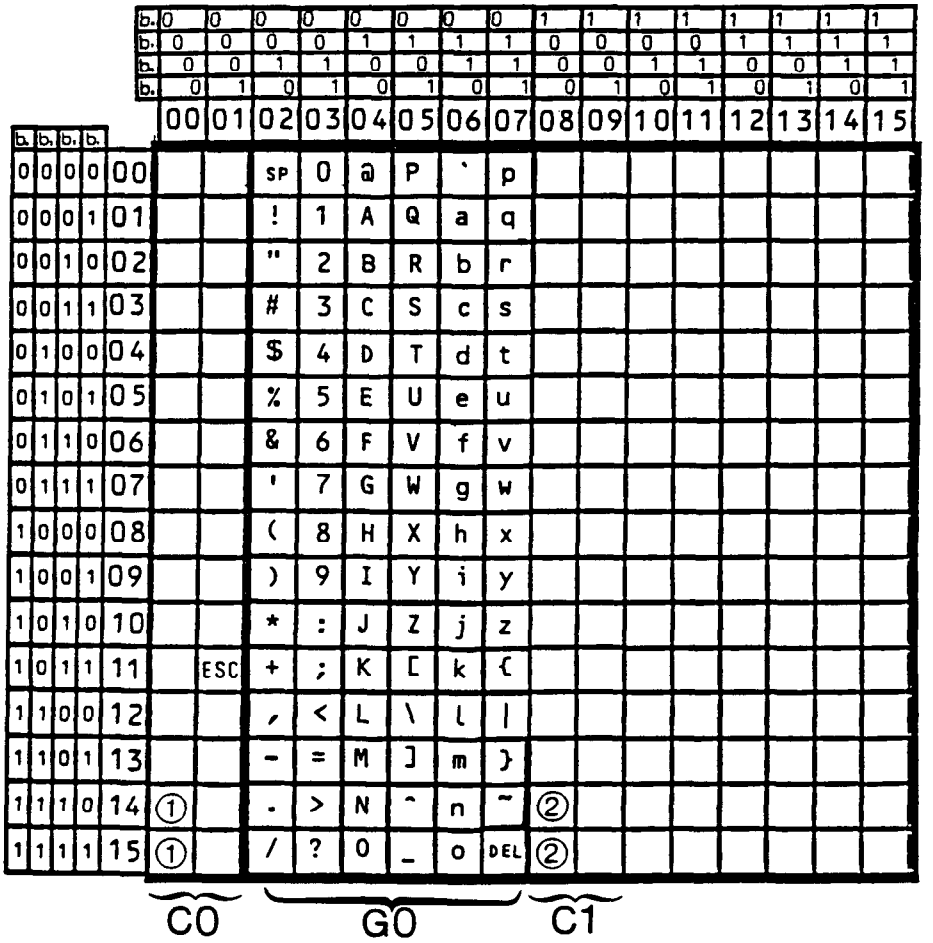
10.4 Switching from one version to another

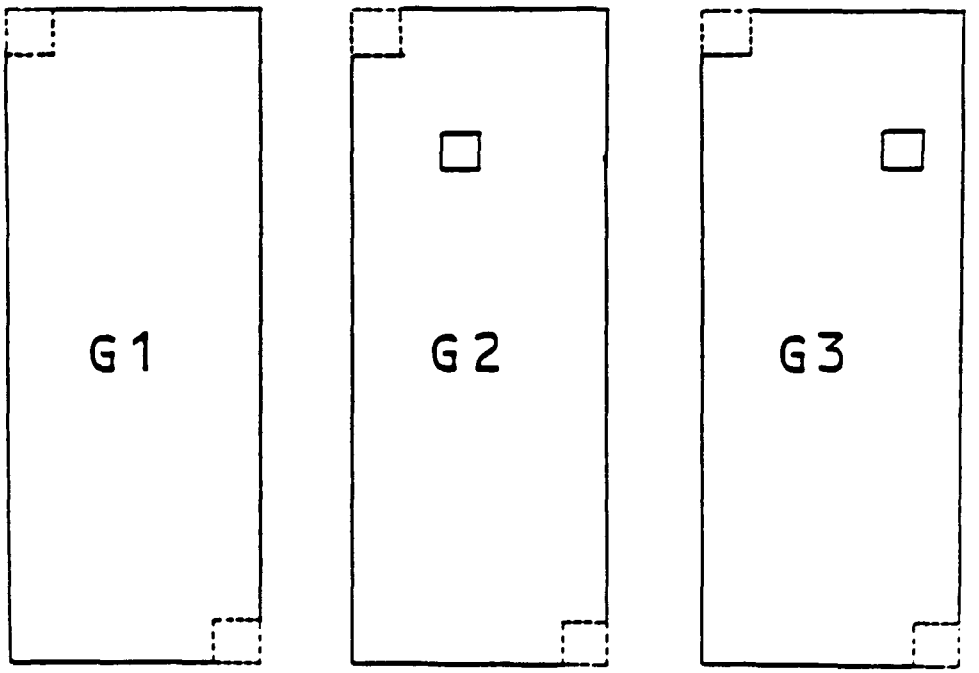
In information interchange, the use of a different version requires that this version be identified. Thus, if any of the C sets or of the G sets is changed, an announcer sequence according to 10.2 and a set of escape sequences according to 10.3 shall be required.

10.5 Switching from one level to another

In information interchange, any change of level, whether or not the C sets and G sets are changed, requires an announcer sequence according to 10.2 and a set of escape sequences according to 10.3.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale





① See 7.1

② See 7.6

Figure 1 - Elements of the 8-bit code

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

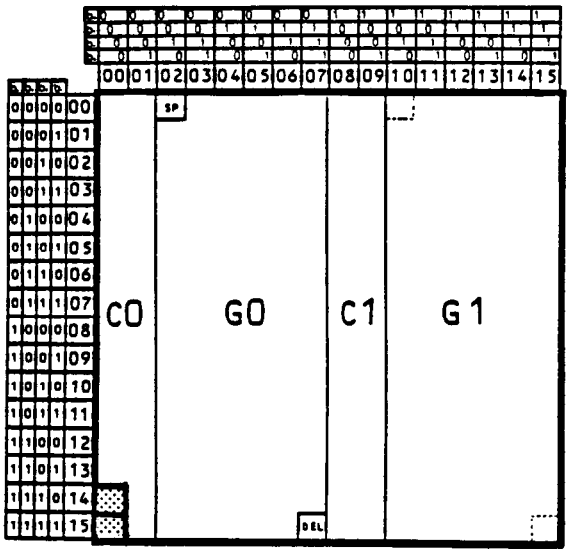


Figure 2 - Level 1

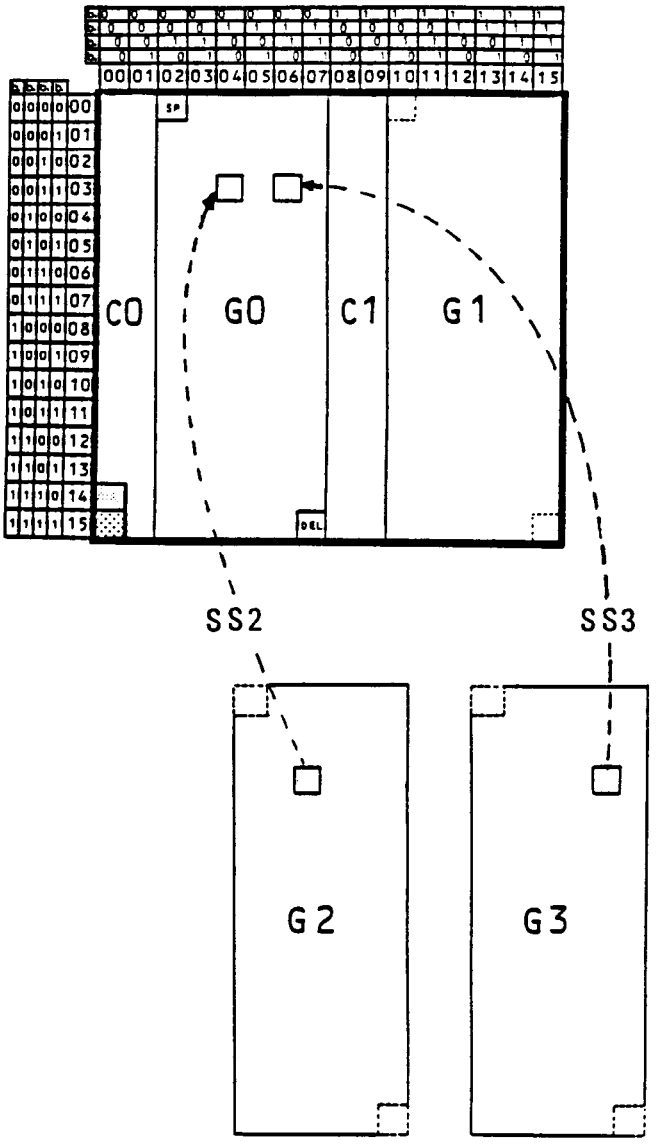


Figure 3 - Level 2

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

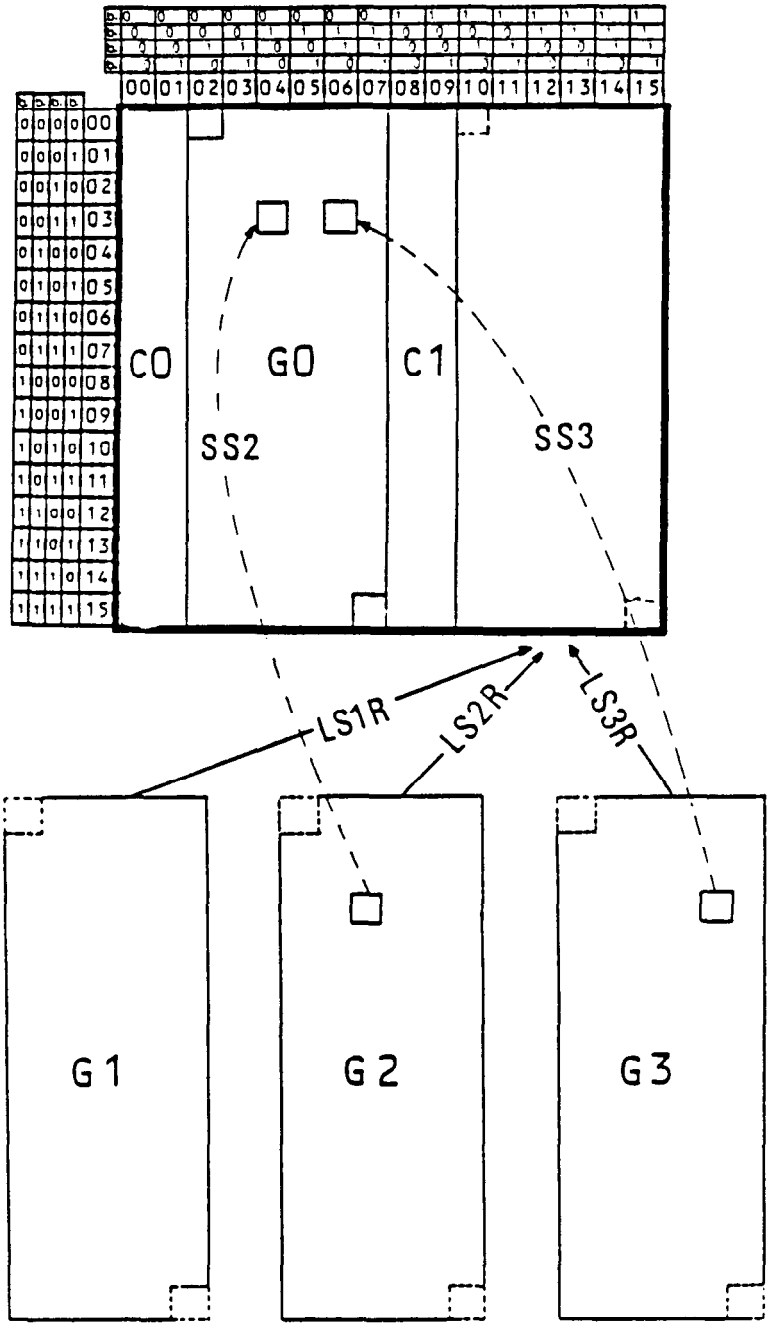


Figure 4 - Level 3

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

Annex A

(normative)

Restrictions applicable to the C0 and C1 sets

The definition of some control functions in this International Standard assumes that data associated with them is to be processed serially in a forward direction. When they are included in strings of data which are processed other than serially in a forward direction or when these control functions are included in data formatted for fixed-record processing they may have undesirable effects or may require additional special treatment to ensure that they result in their desired function.

Whilst this International Standard specifies requirements for bit combinations 00/14, 00/15 and 01/11 of the C0 set, it places the following restrictions on the use of the remaining 29 bit combinations:

a) If control characters described in the C0 set of ISO/IEC 6429 are used, they shall have the coded representations and the definitions specified therein.

b) None of these control characters may be allocated to the C1 set.

c) Transmission control characters are intended to control or facilitate transmission of information over telecommunication networks. Procedures for the use of the transmission control characters on telecommunication networks are the subject of other International Standards, for example ISO 1745.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

Annex B

(informative)

Shift functions

The use of the shift functions mentioned in this International Standard is specified in ISO 2022. They are defined in ISO/IEC 6429 and registered in the International Register of Coded Character Sets to be Used with Escape Sequences.

They are reproduced here for convenience only.

LS1R : ESC 07/14

LS2R : ESC 07/13

LS3R : ESC 07/12

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

Annex C

(informative)

Composite graphic characters

Generally, the specified graphic character set constituting the repertoire of a code is identical with the graphic characters allocated to the bit combinations of that code. In the case of ISO 6937 the graphic character repertoire is larger than the coded graphic character set.

In any case, each graphic character of a graphic character repertoire is represented by a graphic symbol at a character position. After imaging the graphic symbol the active position is advanced by one character position.

By means of the control function GRAPHIC CHARACTER COMBINATION (GCC), specified in ISO 6429, it is possible to represent the graphic symbols of more than one graphic character at one and the same character position. This presentation does not constitute a new graphic symbol but only a juxtaposition of the graphic symbols of the characters considered. This does not create a new coded graphic character nor does it add a graphic character to the specified repertoire. It only images the graphic symbols of characters of that repertoire in a particular manner.

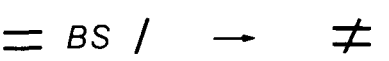
The use of the control functions BACKSPACE (BS) or CARRIAGE RETURN (CR), as permitted by other International Standards, for example by ISO/IEC 646, allows the creation of coded characters in addition to those of the coded character set. Whilst the graphic symbol representing such new coded character consists of a combination of two or more graphic symbols, each representing a character of the coded character set (or of the repertoire where it is larger), this combined graphic symbol must be regarded as a new graphic symbol in its own right and not as a juxtaposition of its components.

Using the latter method makes the graphic character repertoire undefined and requires agreement between the sender and the recipient of the data containing such additional characters. This requirement can often not be met.

This International Standard therefore prohibits the use of BACKSPACE or CARRIAGE RETURN to create additional characters and requires each graphic character to be represented either by a single 8-bit byte or by the defined series of bit combinations as listed exhaustively in ISO/IEC 6937 where the allowed repertoire, which is larger than the coded character set, is fully specified.

Examples

It is not allowed to generate the character "not equal" by combining the characters EQUALS and SOLIDUS by means of the control function BACKSPACE.



It is allowed to generate the presentation of the letters LATIN CAPITAL LETTER P, LATIN SMALL LETTER T and LATIN SMALL LETTER S at one character position by means of the control function GCC.



In the first case a new character with a new graphic symbol has been added to the repertoire. In the second case three characters of the specified repertoire have been imaged in a particular way.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

Annex D

(informative)

Use of bit combinations 00/14 and 00/15

In ISO 2022, the shift functions LS0 and LS1 are represented by the bit combinations 00/14 and 00/15, respectively, in an 8-bit code.

These two bit combinations are not used in this International Standard.

Bit combinations 00/14 and 00/15 are available to represent SO and SI, or LS0 and LS1 when transcoding between a 7-bit code and an 8-bit code. However, this use is outside the scope of this International Standard.

Annex E

(informative)

Main differences between the second edition (1986) and the present (third) edition of this International Standard

E.1 The G0 set is fully defined, i.e. specific graphic characters have been allocated to all 94 character positions.

E.2 A hierarchy of the G sets is specified which allows the assignment of a unique coded representation to a character when that character is present in more than one G set.

E.3 The description of the identification of the C and G sets is separated from the specification of the coding structure.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

UDC 681.3.04:681.327.1

Descriptors: data processing, information interchange, character sets, coded character sets, ISO eight-bit codes, structures, implementation.

Price based on 19 pages

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS

Not for Resale