International Standard 646

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION•МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ•ORGANISATION INTERNATIONALE DE NORMALISATION

Information processing — ISO 7-bit coded character set for information interchange

Traitement de l'information — Jeu ISO de caractères codés à 7 éléments pour l'échange d'information

Second edition — 1983-07-01

UDC 681.3.042 : 003.62 **Ref. No. ISO 646-1983 (E)**

Descriptors : data processing, information interchange, character sets, coded character sets, ISO seven-bit codes, control characters, graphic characters, coded representation.

Price based on 15 pages

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of developing International Standards is carried out through ISO technical committees. Every member body interested in a subject for which a technical committee has been authorized has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work.

Draft International Standards adopted by the technical committees are circulated to the member bodies for approval before their acceptance as International Standards by the ISO Council.

International Standard ISO 646 was developed by Technical Committee ISO/TC 97, *Information processing systems*, and was circulated to the member bodies in March 1982.

It has been approved by the member bodies of the following countries:

Belgium	Germany, F. R.	Romania
Canada	Hungary	South Africa, Rep. of
China	Ireland	Spain
Czechoslovakia	Italy	Sweden
Denmark	Japan	Switzerland
Egypt, Arab Rep. of	Netherlands	United Kingdom
Finland	Norway	USA
France	Poland	

No member body expressed disapproval of the document.

This second edition cancels and replaces the first edition (i.e. ISO 646-1973).

Printed in Switzerland

Contents

Information processing — ISO 7-bit coded character set for information interchange

1 Scope and field of application

1.1 This International Standard specifies a set of 128 characters (control characters and graphic characters such as letters, digits and symbols) with their coded representation. Most of these characters are mandatory and unchangeable, but provision is made for some flexibility to accommodate national and other requirements.

1.2 This International Standard specifies a 7-bit coded character set with a number of options. It also provides guidance on how to exercise the options to define specific national versions and application-orientated versions. Furthermore it specifies the International Reference Version (IRV) in which such options have been exercised.

1.3 This character set is primarily intended for the interchange of information 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 in determining this character set.

1.4 This character set is applicable to all alphabets of Latin letters.

1.5 This character set includes control characters for code extension where its 128 characters are insufficient for particular applications. Procedures for the use of these control characters are specified in ISO 2022.[2]

1.6 The definitions of some control characters in this International Standard assume that data associated with them are 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 they 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.

2 Conformance and implementation

2.1 Conformance

A coded character set is in conformance with this International Standard if it is a version in accordance with clause 6. Equipment claimed to implement this International Standard shall be able to interchange information by means of a version of the 7-bit coded character set, this version shall be identified in any such claim.

2.2 Implementation

The use of this character set requires definitions of its implementation in various media. For example, these could include punched tapes, punched cards, magnetic 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 coded character set in physical media and for transmission, taking into account the need for error checking, is the subject of other International Standards (see the bibliography).

3 Definitions

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

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

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

3.3 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.

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

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

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

3.7 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.

3.8 graphic character : A character, other than a control function, that has a visual representation normally handwritten, printed or displayed.

3.9 position : That part of a code table identified by its column and row co-ordinates.

4 Specification of the coded character set

The bits of the bit combinations of the 7-bit code are identified by b_7, b_6, b_5, b_4, b_3, b_2 and b_1, where b_7 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 127 in binary notation by attributing the following weights to the individual bits :

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

In this International Standard, the bit combinations are identified by notation of the form x/y, where x is a number in the range 0 to 7 and y is a number in the range 0 to 15. The correspondence between the notations of the form x/y and the bit combinations consisting of the bits b_7 to b_1 is as follows :

— x is the number represented by b_7, b_6 and b_5 where these bits are given the weights 4, 2 and 1 respectively;

— y 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 x/y are the same as those used to identify code table positions, where x is the column number and y the row number (see clause 7).

The 128 bit combinations of the 7-bit code represent control characters and graphic characters. The allocation of characters to bit combinations is based on the following principles :

— the bit combinations 0/0 to 1/15 represent 32 control characters;

— the bit combination 2/0 represents the character SPACE, which is interpreted both as a control character and as a graphic character;

— the bit combinations 2/1 to 7/14 represent up to 94 graphic characters as one or more of these bit combinations may be declared to be unused (see 4.3);

— the bit combination 7/15 represents the control character DELETE.

The allocation of individual characters to the bit combinations of the 7-bit code is specified in 4.1, 4.2 and 4.3 below.

This International Standard assigns at least one name to each character. In addition, it specifies an acronym for each control character and for the character SPACE, and a graphic symbol for each graphic character. By convention, only capital letters and hyphens are used for writing the names of the characters, except for small letters. For acronyms only capital letters and digits 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.

4.1 Control characters

The control characters of the 7-bit coded character set are classified in the following categories :

a) Transmission control characters

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 (see the bibliography).

b) Format effectors

Format effectors are mainly intended for the control of the layout and positioning of information on character-imaging devices such as printing and display devices.

c) Code extension control characters

Code extension control characters are used to extend the character set of the code. They may alter the meaning of one or more bit combinations that follow them in the data stream. Procedures for the use of the code extension control characters are specified in ISO 2022.[2]

d) Device control characters

Device control characters are intended for the control of local or remote devices or ancillary devices connected to a data processing or data communication system. These control characters are not intended to control data communication systems; this should be achieved by the use of transmission control characters.

e) Information separators

Information separators are used to separate and qualify data logically. There are four such characters. They may be used either in hierarchical order or non-hierarchically; in the latter case, their specific meanings depend on the application.

f) Other control characters

These are the control characters that fall outside the preceding categories.

The composition of each category, and the allocation of the individual control characters in each category to bit combinations of the 7-bit code are specified in 4.1.1 to 4.1.6. Each of these sub-clauses contains a table consisting of three columns. The first column specifies the acronym of each control character,

the second column specifies the standard name of the control character and the third column, labelled "Coded representation", specifies the bit combination representing the control character concerned.

Detailed functional descriptions of all control characters are given in clause 8.

4.1.1 Transmission control characters

The transmission control characters and their coded representations are specified in table 1.

Table 1 — Transmission control characters — Coded representation

Acronym	Name	Coded representation
SOH	START OF HEADING	0/1
STX	START OF TEXT	0/2
ETX	END OF TEXT	0/3
EOT	END OF TRANSMISSION	0/4
ENQ	ENQUIRY	0/5
ACK	ACKNOWLEDGE	0/6
DLE	DATA LINK ESCAPE	1/0
NAK	NEGATIVE ACKNOWLEDGE	1/5
SYN	SYNCHRONOUS IDLE	1/6
ETB	END OF TRANSMISSION BLOCK	1/7

4.1.2 Format effectors

The format effectors and their coded representations are specified in table 2.

Table 2 — Format effectors — Coded representation

Acronym	Name	Coded representation
BS	BACKSPACE	0/8
HT	HORIZONTAL TABULATION	0/9
LF	LINE FEED	0/10
VT	VERTICAL TABULATION	0/11
FF	FORM FEED	0/12
CR	CARRIAGE RETURN	0/13

4.1.2.1 Concepts

The definitions of the format effectors use the following concepts :

a) A page is composed of a number of lines, each being composed of a number of character positions.

b) Each character position is capable of imaging SPACE or a graphic symbol.

c) The graphic symbol imaged at a character position represents a graphic character, a control function, or a combination of one or more graphic characters and/or control functions.

d) The active position is the character position at which the action required by the next character in the data stream is to be effected. If the next character is a graphic character, it is imaged at that position; if it is a control character, the corresponding function is performed relative to that position.

e) Movements of the active position are effected as follows :

1) The active position is advanced one character position immediately after imaging a SPACE or a graphic character, and upon the execution of the function corresponding to a control character for which a graphic symbol is required to be imaged.

2) The active position is moved to a specified character position upon the execution of the function corresponding to a control character that is defined to cause a movement of the active position (i.e. a format effector).

f) The active position is not moved upon execution of the function corresponding to a control character that is neither required to be imaged by a graphic symbol nor defined to cause a movement of the active position.

g) The effect of an attempt to move the active position beyond the boundaries of a line or a page is not defined by this International Standard.

4.1.2.2 Combined horizontal and vertical movements of the active position

The format effectors are defined for applications in which horizontal and vertical movements of the active position are effected separately. If a single control character is required to effect the action of CARRIAGE RETURN in combination with a vertical movement, the format effector for that vertical movement shall be used. For example, if the function "new line" (equivalent to the combination of CARRIAGE RETURN and LINE FEED) is required as a single control character, bit combination 0/10 shall be used to represent it. This substitution requires agreement between the sender and the recipient of the data, and the format effectors (LINE FEED, VERTICAL TABULATION and/or FORM FEED) that are affected shall be identified (see clause 6).

In order to avoid the need for such prior agreement, to facilitate interchange and to avoid conflicts with specifications in other International Standards, the use of format effectors for vertical movements is deprecated. It is strongly recommended to use two control characters, for example CARRIAGE RETURN (CR) and LINE FEED (LF) to obtain the effect of "new line".

4.1.3 Code extension control characters

The code extension control characters and their coded representations are specified in table 3.

Table 3 — Code extension control characters — Coded representation

Acronym	Name	Coded representation
SO	SHIFT-OUT	0/14
SI	SHIFT-IN	0/15
ESC	ESCAPE	1/11

4.1.4 Device control characters

The device control characters and their coded representations are specified in table 4.

Table 4 — Device control characters — Coded representation

Acronym	Name	Coded representation
DC1	DEVICE CONTROL ONE	1/1
DC2	DEVICE CONTROL TWO	1/2
DC3	DEVICE CONTROL THREE	1/3
DC4	DEVICE CONTROL FOUR	1/4

4.1.5 Information separators

The information separators and their coded representations are specified in table 5.

Table 5 — Information separators — Coded representation

Acronym	Name	Coded representation
IS4 (FS)	INFORMATION SEPARATOR FOUR (FILE SEPARATOR)	1/12
IS3 (GS)	INFORMATION SEPARATOR THREE (GROUP SEPARATOR)	1/13
IS2 (RS)	INFORMATION SEPARATOR TWO (RECORD SEPARATOR)	1/14
IS1 (US)	INFORMATION SEPARATOR ONE (UNIT SEPARATOR)	1/15

Each information separator is given two names. The names INFORMATION SEPARATOR FOUR, INFORMATION SEPARATOR THREE, INFORMATION SEPARATOR TWO and INFORMATION SEPARATOR ONE are the general names. The names FILE SEPARATOR, GROUP SEPARATOR, RECORD SEPARATOR and UNIT SEPARATOR are the specific names and are intended mainly for applications where the information separators are used hierarchically. The ascending order is then US, RS, GS, FS. In this case, data normally delimited by a particular separator cannot be split by a higher-order separator but will be considered as delimited by any higher-order separator.

4.1.6 Other control characters

The control characters outside the categories in 4.1.1 to 4.1.5 and their coded representation, are specified in table 6.

Table 6 — Other control characters — Coded representation

Acronym	Name	Coded representation
NUL	NULL	0/0
BEL	BELL	0/7
CAN	CANCEL	1/8
EM	END OF MEDIUM	1/9
SUB	SUBSTITUTE CHARACTER	1/10
DEL	DELETE	7/15

4.2 Character SPACE

The acronym of the character SPACE is SP and its coded representation is 2/0.

This character is interpreted both as a graphic character and as a control character.

As a graphic character, it has a visual representation consisting of the absence of a graphic symbol.

As a control character, it acts as a format effector that causes the active position to be advanced one character position.

4.3 Graphic characters

The 94 bit combinations 2/1 to 7/14 are used for the representation of graphic characters as specified in 4.3.1, 4.3.2 and 4.3.3 below. Each of the sub-clauses 4.3.1 and 4.3.2 contains a table consisting of three columns. The first column is labelled "Graphic" and specifies the graphic symbol of each graphic character, the second column specifies the standard name of the graphic character and the third column, labelled "Coded representation", specifies the bit combination representing the graphic character concerned.

All graphic characters of any version of the 7-bit coded character set are spacing characters, i.e. they cause the active position to advance.

4.3.1 Unique graphic character allocations

A unique graphic character is allocated to each of the 82 bit combinations 2/1, 2/2, 2/5 to 3/15, 4/1 to 5/10, 5/15 and 6/1 to 7/10. These characters are specified in table 7.

Table 7 — Unique graphic character allocations

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

Graphic	Name	Coded representation
M	CAPITAL LETTER M	4/13
N	CAPITAL LETTER N	4/14
O	CAPITAL LETTER O	4/15
P	CAPITAL LETTER P	5/0
Q	CAPITAL LETTER Q	5/1
R	CAPITAL LETTER R	5/2
S	CAPITAL LETTER S	5/3
T	CAPITAL LETTER T	5/4
U	CAPITAL LETTER U	5/5
V	CAPITAL LETTER V	5/6
W	CAPITAL LETTER W	5/7
X	CAPITAL LETTER X	5/8
Y	CAPITAL LETTER Y	5/9
Z	CAPITAL LETTER Z	5/10
_	LOW LINE, UNDERLINE	5/15
a	SMALL LETTER a	6/1
b	SMALL LETTER b	6/2
c	SMALL LETTER c	6/3
d	SMALL LETTER d	6/4
e	SMALL LETTER e	6/5
f	SMALL LETTER f	6/6
g	SMALL LETTER g	6/7
h	SMALL LETTER h	6/8
i	SMALL LETTER i	6/9
j	SMALL LETTER j	6/10
k	SMALL LETTER k	6/11
l	SMALL LETTER l	6/12
m	SMALL LETTER m	6/13
n	SMALL LETTER n	6/14
o	SMALL LETTER o	6/15
p	SMALL LETTER p	7/0
q	SMALL LETTER q	7/1
r	SMALL LETTER r	7/2
s	SMALL LETTER s	7/3
t	SMALL LETTER t	7/4
u	SMALL LETTER u	7/5
v	SMALL LETTER v	7/6
w	SMALL LETTER w	7/7
x	SMALL LETTER x	7/8
y	SMALL LETTER y	7/9
z	SMALL LETTER z	7/10

4.3.2 Alternative graphic character allocations

Two alternative graphic characters are allocated to each of the bit combinations 2/3 and 2/4. These characters are specified in table 8.

Table 8 — Alternative graphic character allocations

Graphic	Name	Coded representation
£	POUND SIGN	2/3
#	NUMBER SIGN	2/3
$	DOLLAR SIGN	2/4
¤	CURRENCY SIGN	2/4

Either the character POUND SIGN or the character NUMBER SIGN shall be allocated to bit combination 2/3 and either the character DOLLAR SIGN or the character CURRENCY SIGN shall be allocated to bit combination 2/4 (see clause 6).

Unless otherwise agreed between sender and recipient, the graphic symbols £, $ and ¤ do not designate the currency of a specific country.

4.3.3 National or application-orientated graphic character allocations

No specific graphic character is allocated to the ten bit combinations 4/0, 5/11 to 5/14, 6/0, and 7/11 to 7/14. These bit combinations are available for national or application-orientated use. A unique graphic character shall be allocated to each of these bit combinations, or the bit combination shall be declared unused (see clause 6).

5 Composite graphic characters

In any version of the 7-bit coded character set specified according to this International Standard, all graphic characters are spacing characters which cause the active position to move forward. However, by using BACKSPACE or CARRIAGE RETURN, it is possible to image two or more graphic characters at the same character position.

For example, SOLIDUS and EQUALS SIGN can be combined to image "not equals". The character LOW LINE, that may be used as a free-standing character, can also be associated with other character(s) to represent the graphic rendition "underlined".

Diacritical marks may be allocated to the bit combinations specified in 4.3.3 and be available for composing accented letters. For such composition, it is recommended to use a sequence of three characters, the first or last of which is the letter to be accented and the second of which is BACKSPACE. Furthermore, QUOTATION MARK, APOSTROPHE or COMMA can be associated with a letter by means of BACKSPACE for the composition of an accented letter with a diaeresis, an acute accent or a cedilla, respectively.

6 Versions of the coded character set

6.1 General

In order to use the 7-bit coded character set for information interchange, it is necessary to exercise the options left open in clause 4 :

— to each of the bit combinations 2/3 and 2/4 one of the alternative graphic characters specified in 4.3.2 shall be allocated;

— each of the bit combinations 4/0, 5/11 to 5/14, 6/0, and 7/11 to 7/14 shall have a unique graphic character allocated to it, or be declared unused;

— the format effectors, if any, to which the facility of 4.1.2.2 applies, shall be identified.

A graphic character allocated to a bit combination specified in 4.3.1 and 4.3.2 shall not be allocated to any other bit combination. For example the POUND SIGN, if not allocated to bit combination 2/3, shall not be allocated to any other bit combination.

A character set completed in this way is called a "version of ISO 646" (see annex A).

6.2 National versions

6.2.1 The responsibility for defining national versions lies with the national standardization bodies. These bodies shall exercise the options available and make the required selection (see annex A).

6.2.2 If so required, more than one national version can be defined within a country. The different versions shall be separately identified. In particular when for a given bit combination, for example 5/12, alternative graphic characters are required, two different versions shall be identified, even if they differ only by this single character.

6.2.3 If there is in a country no special demand for specific graphic characters, it is strongly recommended that the characters of the International Reference Version (IRV) (see 6.4) be selected and allocated to the same bit combinations as in the IRV.

However, when graphic characters that are different from the characters of the IRV are required, they shall have distinct forms and be given distinctive names which are not in conflict with any of the forms or the names of any of the graphic characters in the IRV.

6.3 Application-orientated versions

Within national or international industries, organizations or professional groups, application-orientated versions can be used. They require precise agreement among the interested parties, who will have to exercise the options available and to make the required selection.

6.4 International Reference Version (IRV)

This version is available for use when there is no requirement to use a national or an application-orientated version. In information interchange, the IRV is assumed unless a particular agreement exists between sender and recipient of the data. The graphic characters allocated to the IRV are specified in table 9.

Table 9 — IRV graphic character allocations

Graphic	Name	Coded representation
#	NUMBER SIGN	2/3
¤	CURRENCY SIGN	2/4
@	COMMERCIAL AT	4/0
[	LEFT SQUARE BRACKET	5/11
\	REVERSE SOLIDUS	5/12
]	RIGHT SQUARE BRACKET	5/13
^	CIRCUMFLEX ACCENT	5/14
`	GRAVE ACCENT	6/0
{	LEFT CURLY BRACKET	7/11
		VERTICAL LINE	7/12
}	RIGHT CURLY BRACKET	7/13
‾	TILDE, OVERLINE	7/14

It should be noted that no substitution is allowed when using the IRV and that the facility of sub-clause 4.1.2.2 does not apply to any format effector.

According to clause 5 it is permitted to use composite graphic characters and there is no limit to their number. Because of this freedom, their processing and imaging may cause difficulties at the receiving end. Therefore agreement between sender and recipient of the data is recommended if composite characters are used.

7 Code tables

A 7-bit code table consists of 128 positions arranged in 8 columns and 16 rows. The columns are numbered 0 to 7, and the rows are numbered 0 to 15.

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

The 128 positions of the code table are in one-to-one correspondence with the bit combinations of the 7-bit code. The notation of a code table position, of the form x/y, is the same as that of the corresponding bit combination (see clause 4).

Each code table position contains a symbol and/or a reference to a clause of this International Standard. When a code table position corresponds to a bit combination that represents a control character or the character SPACE, the symbol is the acronym of the character allocated; otherwise it is the graphic symbol representing the character allocated, if any. A reference to 4.1.2.2, 4.3.2 or 4.3.3 is denoted by ①, ② or ③ respectively.

Table 10 is the basic 7-bit code table. It shows the 7-bit coded character set specified in clause 4 and indicates the options related to format effectors (4.1.2.2), alternative graphic characters (4.3.2) and national or application-orientated use (4.3.3).

Table 11 is the code table for the IRV of the 7-bit coded character set. It shows the result of exercising the three identified options in the manner specified in 6.4.