INTERNATIONAL STANDARD

ISO/IEC 10367

First edition
1991-12-15

Information technology — Standardized coded graphic character sets for use in 8-bit codes

Technologies de l'information — Jeux de caractères graphiques codés normalisés à utiliser dans les codes à un octet

Reference number
ISO/IEC 10367:1991(E)

Contents

Annexes

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 JTC1. 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 10367 was prepared by Joint Technical Committee ISO/IEC JTC1, *Information technology*.

Annexes A and C form an integral part of this International Standard. Annex B is for information only.

Introduction

In the course of the past years two different 8-bit codes were developed by ISO/IEC/JTC1/SC2. An International Standard, ISO/IEC 6937, was developed with a view to satisfying the needs of CCITT-defined Telematic services: Teletex and Videotex. It is based on a primary and supplementary set of graphic characters, the latter containing a series of so-called non-spacing diacritical marks for the generation of accented letters. Thus, the coded representation of a graphic character may consist of one or more bit combinations. The standard specifies the allowed repertoire of 333 characters. ISO/IEC 6937 is applicable not only to 8-bit coding but also to 7-bit coding.

A family of 8-bit code tables, ISO 8859, was also developed to satisfy a need for single-byte coded graphic characters in particular in data processing applications. It consists of several parts, each specifying an 8-bit single-byte coded graphic character set of up to 191 characters. The selection of characters for each set is such that it satisfies the needs of several languages of a large, relatively coherent, geographical area. The different parts of ISO 8859 cover not only the Latin script but also the Arabic, Cyrillic, Greek and Hebrew scripts.

Many applications need a code structure which permits more than one script to be represented in information interchange, for example Latin Alphabet No. 1 ("Western" Europe) with the Greek script or Latin Alphabet No. 2 ("Eastern" Europe) with the Cyrillic script. For this purpose ISO/IEC 4873 is suitable since it allows up to 382 graphic characters grouped in four G sets, by using a small selection of the facilities offered by ISO 2022.

Within this structure the graphic character sets from the various parts of ISO 8859 may be used in conjunction with each other. Since some characters appear in more than one of these sets, rules are needed to avoid violation of the principle of one-to-one relationship between a character and its coded representation when such sets are used together. The new version of ISO/IEC 4873 (3rd edition of 1991) contains such rules.

In a similar way ISO/IEC 6937 may be used in conjunction with additional non-Latin graphic character sets taken from the parts of ISO 8859.

Thus this International Standard specifies a collection of coded graphic character sets usable within the structure of ISO/IEC 4873.

This page intentionally left blank

Information technology — Standardized coded graphic character sets for use in 8-bit codes

1 Scope

This International Standard specifies a unique coded graphic character set for use as G0 set and a series of coded graphic character sets of up to 96 characters for use as the G1, G2 and G3 sets in versions of ISO/IEC 4873. All sets specified in this International Standard are shown as elements of an 8-bit code.

These sets are intended for use in data and text processing applications and may also be used for information interchange. They contain graphic characters used for general purpose applications in typical office environments.

This International Standard does not specify the control functions to be allocated to the C0 and C1 sets of versions of ISO/IEC 4873. ISO/IEC 6429 specifies these control functions; the required control functions shall be selected from that International Standard depending on the application considered.

2 Conformance

2.1 Conformance of information interchange

A coded-character-data-element (CC-data-element) within coded information for interchange is in conformance with this International Standard if all the coded representations of characters within that CC-data-element conform to the requirements of clause 6 and of ISO/IEC 4873.

A claim of conformance shall identify the selected character sets.

2.2 Conformance of devices

A device is in conformance with this International Standard if it conforms to the requirements of 2.2.1, and either or both of 2.2.2 and 2.2.3. A claim of conformance shall identify the document which contains the description specified in 2.2.1, and shall identify the selected character sets.

2.2.1 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 2.2.2 and 2.2.3.

2.2.2 Originating devices

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

2.2.3 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, and shall make the corresponding characters available to its user in such a way that the user can identify them from among those from the selected character sets, and can distinguish them from each other.

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

ISO/IEC 4873:1991, *Information technology - 8-bit code for information interchange - Structure and rules for implementation.*

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

ISO/IEC 6937: [1] , *Information technology - Coded graphic character set for the communication of texts using the Latin alphabet.*

ISO/IEC 8859, *Information processing - 8-bit single-byte coded graphic character sets.*

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

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

4.3 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.

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.

4.4 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 coded representation by one or more bit combinations.

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

4.6 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.

4.7 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 a gateway function.

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

1) To be published.

4.9 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.

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

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

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

4.13 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 2.2 above is satisfied.

5 Notation, code tables and names

Each coded character set is represented as a part of an 8-bit code table. These sets are identical with those of the International Register where they are shown in a 7-bit environment.

5.1 Notation

The bits of the bit combination of the 8-bit code are identified by b8, b7, b6, b5, b4, b3, b2 and b1, where b8 is the highest-order, or most-significant bit and b1 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.

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

Using these weights, these 8-bit combinations represent numbers in the range 0 to 255.

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 b8 to b1 is as follows:

- xx is the number represented by b8, b7, b6 and b5 where these bits are given the weights 8, 4, 2 and 1 respectively;
- yy is the number represented by b4, b3, b2 and b1 where these bits are given the weights 8, 4, 2 and 1 respectively.

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

5.2 Layout of the code tables

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

The Basic G0 Set is shown in columns 02 to 07 of an 8-bit code table.

All other G sets are shown in columns 10 to 15 of an 8-bit code table.

5.3 Names

This International Standard assigns a unique name to each character. In addition, it specifies an acronym for the characters NO-BREAK-SPACE and SOFT HYPHEN, and a graphic symbol for each graphic character. By convention only capital letters, space and hyphen are used for writing the names of the characters. 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, except for NO-BREAK SPACE (NBSP) and SOFT HYPHEN (SHY), 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 (see annex B).

5.3.1 NO-BREAK SPACE (NBSP)

A graphic character the visual representation of which consists of the absence of a graphic symbol, for use when a line break is to be prevented in the text as presented.

5.3.2 SOFT HYPHEN (SHY)

A graphic character that is imaged by a graphic symbol identical with, or similar to, that representing HYPHEN, for use when a line break has been established within a word.

6 Specification of the character sets

Each character set, except the Supplementary Set of ISO/IEC 6937, is specified by a code table according to 5.2, and by the list of the names of the characters. For each set its name, registration number, escape sequences and number of characters are indicated. In addition, the Field of Utilisation indicates the languages, the alphabets of which can be represented by the combination of the Basic G0 Set with the set considered.

The character SPACE is deemed to be available with any character set and, therefore, is not specified with a character set selected from this International Standard.

The use of the Supplementary Set of ISO/IEC 6937 is separately specified in annex A.

CODE TABLES
AND
LISTS OF CHARACTER NAMES

Name : BASIC G0 SET
Registration : No. 6
Escape sequence : ESC 02/08 04/02
Number of characters : 94

Field of Utilisation

The Basic G0 Set is intended for use with one or more other sets of this Repertoire in versions of ISO 4873.

Table 1 - Basic G0 Set

b4	b3	b2	b1		02	03	04	05	06	07
0	0	0	0	00		0	@	P	`	p
0	0	0	1	01	!	1	A	Q	a	q
0	0	1	0	02	"	2	B	R	b	r
0	0	1	1	03	#	3	C	S	c	s
0	1	0	0	04	$	4	D	T	d	t
0	1	0	1	05	%	5	E	U	e	u
0	1	1	0	06	&	6	F	V	f	v
0	1	1	1	07	'	7	G	W	g	w
1	0	0	0	08	(	8	H	X	h	x
1	0	0	1	09	)	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	

Table 2 - Names of the characters of table 1

Bit comb.	Name
02/01	EXCLAMATION MARK
02/02	QUOTATION MARK
02/03	NUMBER SIGN
02/04	DOLLAR SIGN
02/05	PERCENT SIGN
02/06	AMPERSAND
02/07	APOSTROPHE
02/08	LEFT PARENTHESIS
02/09	RIGHT PARENTHESIS
02/10	ASTERISK
02/11	PLUS SIGN
02/12	COMMA
02/13	HYPHEN-MINUS
02/14	FULL STOP
02/15	SOLIDUS
03/00	DIGIT ZERO
03/01	DIGIT ONE
03/02	DIGIT TWO
03/03	DIGIT THREE
03/04	DIGIT FOUR
03/05	DIGIT FIVE
03/06	DIGIT SIX
03/07	DIGIT SEVEN
03/08	DIGIT EIGHT
03/09	DIGIT NINE
03/10	COLON
03/11	SEMICOLON
03/12	LESS-THAN SIGN
03/13	EQUALS SIGN
03/14	GREATER-THAN SIGN
03/15	QUESTION MARK
04/00	COMMERCIAL AT
04/01	LATIN CAPITAL LETTER A
04/02	LATIN CAPITAL LETTER B
04/03	LATIN CAPITAL LETTER C
04/04	LATIN CAPITAL LETTER D
04/05	LATIN CAPITAL LETTER E
04/06	LATIN CAPITAL LETTER F
04/07	LATIN CAPITAL LETTER G
04/08	LATIN CAPITAL LETTER H
04/09	LATIN CAPITAL LETTER I
04/10	LATIN CAPITAL LETTER J
04/11	LATIN CAPITAL LETTER K
04/12	LATIN CAPITAL LETTER L
04/13	LATIN CAPITAL LETTER M
04/14	LATIN CAPITAL LETTER N
04/15	LATIN CAPITAL LETTER O

Bit comb.	Name
05/00	LATIN CAPITAL LETTER P
05/01	LATIN CAPITAL LETTER Q
05/02	LATIN CAPITAL LETTER R
05/03	LATIN CAPITAL LETTER S
05/04	LATIN CAPITAL LETTER T
05/05	LATIN CAPITAL LETTER U
05/06	LATIN CAPITAL LETTER V
05/07	LATIN CAPITAL LETTER W
05/08	LATIN CAPITAL LETTER X
05/09	LATIN CAPITAL LETTER Y
05/10	LATIN CAPITAL LETTER Z
05/11	LEFT SQUARE BRACKET
05/12	REVERSE SOLIDUS
05/13	RIGHT SQUARE BRACKET
05/14	CIRCUMFLEX ACCENT
05/15	LOW LINE
06/00	GRAVE ACCENT
06/01	LATIN SMALL LETTER A
06/02	LATIN SMALL LETTER B
06/03	LATIN SMALL LETTER C
06/04	LATIN SMALL LETTER D
06/05	LATIN SMALL LETTER E
06/06	LATIN SMALL LETTER F
06/07	LATIN SMALL LETTER G
06/08	LATIN SMALL LETTER H
06/09	LATIN SMALL LETTER I
06/10	LATIN SMALL LETTER J
06/11	LATIN SMALL LETTER K
06/12	LATIN SMALL LETTER L
06/13	LATIN SMALL LETTER M
06/14	LATIN SMALL LETTER N
06/15	LATIN SMALL LETTER O
07/00	LATIN SMALL LETTER P
07/01	LATIN SMALL LETTER Q
07/02	LATIN SMALL LETTER R
07/03	LATIN SMALL LETTER S
07/04	LATIN SMALL LETTER T
07/05	LATIN SMALL LETTER U
07/06	LATIN SMALL LETTER V
07/07	LATIN SMALL LETTER W
07/08	LATIN SMALL LETTER X
07/09	LATIN SMALL LETTER Y
07/10	LATIN SMALL LETTER Z
07/11	LEFT CURLY BRACKET
07/12	VERTICAL LINE
07/13	RIGHT CURLY BRACKET
07/14	TILDE

Name : LATIN ALPHABET No. 1, SUPPLEMENTARY SET
Registration : No. 100
Escape sequences : G1 : ESC 02/13 04/01
: G2 : ESC 02/14 04/01
: G3 : ESC 02/15 04/01
Number of characters : 96

Field of Utilisation

Danish, Dutch, English, Faroese, Finnish, German, Icelandic, Irish, Italian, Norwegian, Portuguese, Spanish and Swedish.

Table 3 - Latin Alphabet No. 1, Supplementary Set

b8	b7	b6	b5	10	11	12	13	14	15
				b4 1	1	1	1	1	1
				b3 0	0	1	1	1	1
				b2 1	1	0	0	1	1
				b1 0	1	0	1	0	1
0	0	0	0	00	NBSP	°	À	Ð	à	ð
0	0	0	1	01	¡	±	Á	Ñ	á	ñ
0	0	1	0	02	¢	²	Â	Ò	â	ò
0	0	1	1	03	£	³	Ã	Ó	ã	ó
0	1	0	0	04	¤	´	Ä	Ô	ä	ô
0	1	0	1	05	¥	µ	Å	Õ	å	õ
0	1	1	0	06	¦	¶	Æ	Ö	æ	ö
0	1	1	1	07	§	·	Ç	×	ç	÷
1	0	0	0	08	¨	¸	È	Ø	è	ø
1	0	0	1	09	©	¹	É	Ù	é	ù
1	0	1	0	10	ª	º	Ê	Ú	ê	ú
1	0	1	1	11	«	»	Ë	Û	ë	û
1	1	0	0	12	¬	¼	Ì	Ü	ì	ü
1	1	0	1	13	SHY	½	Í	Ý	í	ý
1	1	1	0	14	®	¾	Î	Þ	î	þ
1	1	1	1	15	¯	¿	Ï	ß	ï	ÿ

Table 4 - Names of the characters of table 3

Bit comb.	Name
10/00	NO-BREAK SPACE
10/01	INVERTED EXCLAMATION MARK
10/02	CENT SIGN
10/03	POUND SIGN
10/04	CURRENCY SIGN
10/05	YEN SIGN
10/06	BROKEN BAR
10/07	SECTION SIGN
10/08	DIAERESIS
10/09	COPYRIGHT SIGN
10/10	FEMININE ORDINAL INDICATOR
10/11	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
10/12	NOT SIGN
10/13	SOFT HYPHEN
10/14	REGISTERED SIGN
10/15	MACRON
11/00	DEGREE SIGN
11/01	PLUS-MINUS SIGN
11/02	SUPERSCRIPT TWO
11/03	SUPERSCRIPT THREE
11/04	ACUTE ACCENT
11/05	MICRO SIGN
11/06	PILCROW SIGN
11/07	MIDDLE DOT
11/08	CEDILLA
11/09	SUPERSCRIPT ONE
11/10	MASCULINE ORDINAL INDICATOR
11/11	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
11/12	VULGAR FRACTION ONE QUARTER
11/13	VULGAR FRACTION ONE HALF
11/14	VULGAR FRACTION THREE QUARTERS
11/15	INVERTED QUESTION MARK
12/00	LATIN CAPITAL LETTER A WITH GRAVE
12/01	LATIN CAPITAL LETTER A WITH ACUTE
12/02	LATIN CAPITAL LETTER A WITH CIRCUMFLEX
12/03	LATIN CAPITAL LETTER A WITH TILDE
12/04	LATIN CAPITAL LETTER A WITH DIAERESIS
12/05	LATIN CAPITAL LETTER A WITH RING ABOVE
12/06	LATIN CAPITAL LIGATURE AE
12/07	LATIN CAPITAL LETTER C WITH CEDILLA
12/08	LATIN CAPITAL LETTER E WITH GRAVE
12/09	LATIN CAPITAL LETTER E WITH ACUTE
12/10	LATIN CAPITAL LETTER E WITH CIRCUMFLEX
12/11	LATIN CAPITAL LETTER E WITH DIAERESIS
12/12	LATIN CAPITAL LETTER I WITH GRAVE
12/13	LATIN CAPITAL LETTER I WITH ACUTE
12/14	LATIN CAPITAL LETTER I WITH CIRCUMFLEX
12/15	LATIN CAPITAL LETTER I WITH DIAERESIS

Bit comb.	Name
13/00	LATIN CAPITAL LETTER ETH (Icelandic)
13/01	LATIN CAPITAL LETTER N WITH TILDE
13/02	LATIN CAPITAL LETTER O WITH GRAVE
13/03	LATIN CAPITAL LETTER O WITH ACUTE
13/04	LATIN CAPITAL LETTER O WITH CIRCUMFLEX
13/05	LATIN CAPITAL LETTER O WITH TILDE
13/06	LATIN CAPITAL LETTER O WITH DIAERESIS
13/07	MULTIPLICATION SIGN
13/08	LATIN CAPITAL LETTER O WITH STROKE
13/09	LATIN CAPITAL LETTER U WITH GRAVE
13/10	LATIN CAPITAL LETTER U WITH ACUTE
13/11	LATIN CAPITAL LETTER U WITH CIRCUMFLEX
13/12	LATIN CAPITAL LETTER U WITH DIAERESIS
13/13	LATIN CAPITAL LETTER Y WITH ACUTE
13/14	LATIN CAPITAL LETTER THORN (Icelandic)
13/15	LATIN SMALL LETTER SHARP S (German)
14/00	LATIN SMALL LETTER A WITH GRAVE
14/01	LATIN SMALL LETTER A WITH ACUTE
14/02	LATIN SMALL LETTER A WITH CIRCUMFLEX
14/03	LATIN SMALL LETTER A WITH TILDE
14/04	LATIN SMALL LETTER A WITH DIAERESIS
14/05	LATIN SMALL LETTER A WITH RING ABOVE
14/06	LATIN SMALL LIGATURE AE
14/07	LATIN SMALL LETTER C WITH CEDILLA
14/08	LATIN SMALL LETTER E WITH GRAVE
14/09	LATIN SMALL LETTER E WITH ACUTE
14/10	LATIN SMALL LETTER E WITH CIRCUMFLEX
14/11	LATIN SMALL LETTER E WITH DIAERESIS
14/12	LATIN SMALL LETTER I WITH GRAVE
14/13	LATIN SMALL LETTER I WITH ACUTE
14/14	LATIN SMALL LETTER I WITH CIRCUMFLEX
14/15	LATIN SMALL LETTER I WITH DIAERESIS
15/00	LATIN SMALL LETTER ETH (Icelandic)
15/01	LATIN SMALL LETTER N WITH TILDE
15/02	LATIN SMALL LETTER O WITH GRAVE
15/03	LATIN SMALL LETTER O WITH ACUTE
15/04	LATIN SMALL LETTER O WITH CIRCUMFLEX
15/05	LATIN SMALL LETTER O WITH TILDE
15/06	LATIN SMALL LETTER O WITH DIAERESIS
15/07	DIVISION SIGN
15/08	LATIN SMALL LETTER O WITH STROKE
15/09	LATIN SMALL LETTER U WITH GRAVE
15/10	LATIN SMALL LETTER U WITH ACUTE
15/11	LATIN SMALL LETTER U WITH CIRCUMFLEX
15/12	LATIN SMALL LETTER U WITH DIAERESIS
15/13	LATIN SMALL LETTER Y WITH ACUTE
15/14	LATIN SMALL LETTER THORN (Icelandic)
15/15	LATIN SMALL LETTER Y WITH DIAERESIS