

INTERNATIONAL STANDARD

ISO
9036

First edition
1987-04-15



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

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

Jeu de caractères arabes codés à 7 éléments pour l'échange d'information

STANDARDSISO.COM : Click to view the full PDF of ISO 9036:1987

Reference number
ISO 9036 : 1987 (E)

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established 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. They are approved in accordance with ISO procedures requiring at least 75 % approval by the member bodies voting.

International Standard ISO 9036 was prepared by Technical Committee ISO/TC 97, *Information processing systems*.

Users should note that all International Standards undergo revision from time to time and that any reference made herein to any other International Standard implies its latest edition, unless otherwise stated.

STANDARDSISO.COM : Click to view the full PDF of ISO 9036:1987

Contents	Page
1 Scope and field of application	1
2 Conformance and implementation	1
2.1 Conformance	1
2.2 Implementation	1
3 References	1
4 Definitions	1
4.1 bit combination	1
4.2 character	1
4.3 coded character set; code	1
4.4 code extension	1
4.5 code table	1
4.6 control character	1
4.7 control function	1
4.8 graphic character	2
4.9 graphic symbol	2
4.10 position	2
5 Specification of the coded character set	2
5.1 Control characters	2
5.2 Character SPACE	4
5.3 Graphic characters	4
6 Composite graphic characters	7
7 Vowel characters	7
8 Designation of the character set of this International Standard	7
9 Code table	7
10 Description of the control characters	9
10.1 ACK ACKNOWLEDGE	9
10.2 BEL BELL	9
10.3 BS BACKSPACE	9
10.4 CAN CANCEL	9
10.5 CR CARRIAGE RETURN	9
10.6 DC1 DEVICE CONTROL ONE	9
10.7 DC2 DEVICE CONTROL TWO	9
10.8 DC3 DEVICE CONTROL THREE	9
10.9 DC4 DEVICE CONTROL FOUR	9
10.10 DEL DELETE	9
10.11 DLE DATA LINK ESCAPE	9

10.12	EM	END OF MEDIUM	9
10.13	ENQ	ENQUIRY	9
10.14	EOT	END OF TRANSMISSION	9
10.15	ESC	ESCAPE	9
10.16	ETB	END OF TRANSMISSION BLOCK	9
10.17	ETX	END OF TEXT	9
10.18	FF	FORM FEED	9
10.19	HT	HORIZONTAL TABULATION	10
10.20	IS1 (US)	INFORMATION SEPARATOR ONE (UNIT SEPARATOR) ..	10
10.21	IS2 (RS)	INFORMATION SEPARATOR TWO (RECORD SEPARATOR)	10
10.22	IS3 (GS)	INFORMATION SEPARATOR THREE (GROUP SEPARATOR)	10
10.23	IS4 (FS)	INFORMATION SEPARATOR FOUR (FILE SEPARATOR) ..	10
10.24	LF	LINE FEED	10
10.25	NACK	NEGATIVE ACKNOWLEDGE	10
10.26	NUL	NULL	10
10.27	SI	SHIFT-IN	10
10.28	SO	SHIFT-OUT	10
10.29	SOH	START OF HEADING	10
10.30	STX	START OF TEXT	10
10.31	SUB	SUBSTITUTE CHARACTER	10
10.32	SYN	SYNCHRONOUS IDLE	10
10.33	VT	VERTICAL TABULATION	10
11		Peculiarities relating to the conventional direction of Arabic writing	10

STANDARDSISO.COM : Click to view the full PDF of ISO 9036:1987

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

1 Scope and field of application

1.1 This International Standard specifies a set of 120 characters (control characters and graphic characters such as letters, digits and symbols) with their coded representation. These characters are mandatory and unchangeable.

1.2 This character set is primarily intended for the interchange of information using the Arabic language 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.

1.3 This character set includes control characters for code extension where its 120 characters are insufficient for particular applications. Procedures for the use of these control characters are specified in ISO 2022.

1.4 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 comprises the characters listed in clause 5 with the specified coded representation.

Equipment claimed to implement this International Standard shall be able to interchange information by means of this 7-bit coded character set.

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.

3 References

ISO 646, *Information processing — ISO 7-bit coded character set for information interchange*.

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

Arab Standard ASMO 449.

International Register of Coded Character Sets to Be Used with Escape Sequences.

4 Definitions

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

4.1 bit combination: An ordered set of bits that represents a character or is used as part of the representation of a character.

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

4.3 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.

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

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

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

4.7 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.

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

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

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

5 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 86 graphic characters; eight of these bit combinations shall not be used;
- 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 clause 5.

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, apostrophes and hyphens 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, 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 symbols representing the characters when they are imaged.

5.1 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.

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.

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 5.1.1 to 5.1.6; 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 specifies the bit combination representing the control character concerned.

Detailed functional descriptions of all control characters are given in clause 10.

5.1.1 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

5.1.2 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

5.1.2.1 Concepts

The definitions of the format effectors use the following concepts:

- A page is composed of a number of lines, each being composed of a number of character positions.
- 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, its graphic symbol 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 symbol, 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.

5.1.2.2 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 10).

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 movement to effect combined horizontal and 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".

5.1.3 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

5.1.4 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

5.1.2 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	INFORMATION SEPARATOR FOUR (FILE SEPARATOR)	1/12
IS3	INFORMATION SEPARATOR THREE (GROUP SEPARATOR)	1/13
IS2	INFORMATION SEPARATOR TWO (RECORD SEPARATOR)	1/14
IS1	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.

5.1.6 Other control characters

The control characters outside the categories in 5.1.3 to 5.1.5 and their coded representations 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

5.2 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.

5.3 Graphic characters

From the 94 bit combinations 2/1 to 7/14 there are 86 used for the representation of graphic characters as specified below.

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

A unique graphic character is allocated to each of the 86 bit combinations 2/1 to 7/2 and 7/11 to 7/14 (see table 7).

Bit combinations 7/3 to 7/10 shall not be used.

Table 7 — Unique graphic character allocations

Graphic symbol	Name	Coded representation
!	EXCLAMATION MARK	2/1
"	QUOTATION MARK	2/2
#	NUMBER SIGN	2/3
¤	CURRENCY SIGN	2/4
%	PERCENT SIGN	2/5
&	AMPERSAND	2/6

Table 7 – Unique graphic character allocations (continued)

Graphic symbol	Name	Coded representation	Graphic symbol	Name	Coded representation
'	APOSTROPHE	2/7	؛	ARABIC SEMICOLON	3/11
)	RIGHT PARENTHESIS	2/8	<	ARABIC LESS-THAN SIGN	3/12
(LEFT PARENTHESIS	2/9	=	EQUALS SIGN	3/13
*	ASTERISK	2/10	>	ARABIC GREATER-THAN SIGN	3/14
+	PLUS SIGN	2/11	؟	ARABIC QUESTION MARK	3/15
‘	ARABIC COMMA	2/12	@	COMMERCIAL AT	4/0
-	HYPHEN, MINUS SIGN	2/13	ء	HAMZAH	4/1
.	FULL STOP	2/14	أ	MADDAH ON ALEF	4/2
/	SOLIDUS	2/15	إ	HAMZAH ON ALEF	4/3
0	DIGIT ZERO	3/0	آ	HAMZAH ON WAW	4/4
1	DIGIT ONE	3/1	أ	HAMZAH UNDER ALEF	4/5
2	DIGIT TWO	3/2	أ	HAMZAH ON YA	4/6
3	DIGIT THREE	3/3	ا	ALEF	4/7
4	DIGIT FOUR	3/4	ب	BA'A	4/8
5	DIGIT FIVE	3/5	ة	TA'A MARBUTA	4/9
6	DIGIT SIX	3/6	ت	TA'A	4/10
7	DIGIT SEVEN	3/7	ث	THA'A	4/11
8	DIGIT EIGHT	3/8	ج	JEEM	4/12
9	DIGIT NINE	3/9	ح	HA'A	4/13
:	COLON	3/10	خ	KHA'A	4/14

Table 7 — Unique graphic character allocations (concluded)

Graphic symbol	Name	Coded representation
ﺩ	DAL	4/15
ﺛ	THAL	5/0
ﺭ	RA	5/1
ﺯ	ZAIN	5/2
ﺲ	SEEN	5/3
ﺶ	SHEEN	5/4
ﺲ	SAD	5/5
ﺪ	DAD	5/6
ﺗ	TAH	5/7
ﺪﻩ	DHAH	5/8
ﺀ	AIN	5/9
ﻏ	GHAIN	5/10
]	RIGHT SQUARE BRACKET	5/11
/	REVERSE SOLIDUS	5/12
[LEFT SQUARE BRACKET	5/13
^	UPWARD ARROW HEAD	5/14
_	UNDERLINE	5/15
	TATWEEL	6/0
ﻑ	FA	6/1
ﻗ	QAF	6/2

Graphic symbol	Name	Coded representation
ﻙ	CAF	6/3
ﻝ	LAM	6/4
ﻡ	MEEM	6/5
ﻥ	NOON	6/6
ﻩ	HA	6/7
ﻭ	WAW	6/8
ﻱ	ALEF MAQSURA	6/9
ﻱ	YA	6/10
”	FAT’HATAN	6/11
”	DAMMATAN	6/12
”	KASRATAN	6/13
’	FAT’HA	6/14
ﻩ	DAMMAH	6/15
’	KASRAH	7/0
ﻩ	SHADDAH	7/1
ﻩ	SUKUN	7/2
}	RIGHT CURLY BRACKET	7/11
	VERTICAL LINE	7/12
{	LEFT CURLY BRACKET	7/13
—	OVERLINE	7/14

6 Composite graphic characters

In the 7-bit coded character set specified by 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 UNDERLINE, that may be used as a free-standing character, can also be associated with other character(s) to represent the graphic rendition "underlined".

7 Vowel characters

No graphic symbol may be used for the composition of Arabic vowelized letters. The characters allocated to positions 6/11 to 7/2 are vowel characters and not graphic vowel symbols to be combined with other Arabic letters, thus the character set of this International Standard does not contain diacritical marks.

8 Designation of the character set of this International Standard

The set of graphic characters of this International Standard has been registered in the International Register of Coded

Character Sets to Be Used with Escape Sequences. The allocated escape sequences are

Designation as a G0 set: ESC 2/8 6/11
 Designation as a G1 set: ESC 2/9 6/11
 Designation as a G2 set: ESC 2/10 6/11
 Designation as a G3 set: ESC 2/11 6/11

9 Code table

The 7-bit code table (see table 8) 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 5).

Each code table position contains a symbol or is shaded for those positions which shall not be used. 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.

STANDARDSISO.COM : Click to view the PDF of ISO 9036:1987

Table 8 — 7-bit code table

					b7	0	0	0	0	1	1	1	1
					b6	0	0	1	1	0	0	1	1
					b5	0	1	0	1	0	1	0	1
						0	1	2	3	4	5	6	7
b4	b3	b2	b1										
0	0	0	0	0	NUL	DLE	SP	0	ا	ذ	—		
0	0	0	1	1	SOH	DC1	!	1	ء	ر	ف		
0	0	1	0	2	STX	DC2	"	2	آ	ز	ق		
0	0	1	1	3	ETX	DC3	#	3	أ	س	ك		
0	1	0	0	4	EOT	DC4	α	4	و	ش	ل		
0	1	0	1	5	ENQ	NAK	%	5	ا	ص	م		
0	1	1	0	6	ACK	SYN	&	6	ر	ض	ن		
0	1	1	1	7	BEL	ETB	'	7	ا	ط	ه		
1	0	0	0	8	BS	CAN)	8	ب	ظ	و		
1	0	0	1	9	HT	EM	(9	ة	ع	ى		
1	0	1	0	10	LF	SUB	*	:	ت	غ	ي		
1	0	1	1	11	VT	ESC	+	؛	ث]	=	}	
1	1	0	0	12	FF	IS4	,	>	ج	\	"		
1	1	0	1	13	CR	IS3	-	=	ح	[=	{	
1	1	1	0	14	SO	IS2	.	<	خ	^	'	-	
1	1	1	1	15	SI	IS1	/	؟	د	_	'	DEL	