

L2/01-239

INTERNATIONAL
STANDARD

ISO
11822

First edition
1996-12-15

**Information and documentation —
Extension of the Arabic alphabet coded
character set for bibliographic information
interchange**

*Information et documentation — Extension du jeu de caractères codés de
l'alphabet arabe pour les échanges d'informations bibliographiques*



Reference number
ISO 11822:1996(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. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

International Standard ISO 11822 was prepared by Technical Committee ISO/TC 46, *Information and documentation*, Subcommittee SC 4, *Computer applications in information and documentation*.

Annexes A and B of this International Standard are for information only.

© ISO 1996

All rights reserved. Unless otherwise specified, 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.

International Organization for Standardization
Case Postale 56 • CH-1211 Genève 20 • Switzerland

Printed in Switzerland

Information and documentation — Extension of the Arabic alphabet coded character set for bibliographic information interchange

1 Scope

1.1 This International Standard specifies a set of 90 graphic characters with their coded representations. It consists of a code table and a legend showing character codes, graphics and character names. Explanatory notes are also included. The character set is primarily intended for the interchange of information among data processing systems and within message transmission systems.

1.2 These characters, together with characters in the international reference version of ISO 9036, constitute a character set for the international interchange of bibliographic citations, including their annotations, in the Arabic script. The sets may be used in a 7-bit or an 8-bit environment in accordance with ISO/IEC 2022.

1.3 This character set, with characters from ISO 9036 (see annex A), is intended for information in the following languages:

Adighe	Farsi	Malay
Arabic	Hausa	Moplah
Avaric	Kashmiri	Pushto
Baluchi	Kirghiz	Sindhi
Berber	Kurdish	Turkish
Coptic	Lahnda	Uighur
Dargwa	Lak	Urdu

1.4 The graphic representation of characters defined in this International Standard are given in their isolated forms only. Initial, medial, and final forms, as well as special presentation forms which occur in ligatures are not within the scope of this International Standard.

2 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 indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO/IEC 2022:1994, *Information technology — Character code structure and extension techniques*.

ISO 9036:1987, *Information processing — Arabic 7-bit coded character set for information interchange*.

*International register of character sets to be identified by means of escape sequences.*¹⁾

¹⁾ Available on application to the Secretariat of the Registration Authority: ECMA, 114 rue du Rhône, CH-1204 Genève, Switzerland.

3 Implementation

3.1 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 annex B).

3.2 The implementation of this International Standard is in accordance with the provisions of ISO/IEC 2022²⁾ and is identified by an escape sequence. (To be assigned.)

3.3 The unassigned positions in the code table shall not be utilized in the international interchange of bibliographic information.

2) G0: ESC 2/8 F; G1: ESC 2/9 F; G2: ESC 2/10 F; G3: ESC 2/11 F ("F" represents the final character of the escape sequence).

4 Code table for extended Arabic coded characters

Table 1 is the code table for extended Arabic coded characters.

Table 1

					b ₇	0	0	0	0	1	1	1	1
					b ₆	0	0	1	1	0	0	1	1
					b ₅	0	1	0	1	0	1	0	1
b ₄	b ₃	b ₂	b ₁		0	1	2	3	4	5	6	7	
0	0	0	0	0				خ	ر	ط	ك	و	
0	0	0	1	1			أ	ح	ر	ط	ك	و	
0	0	1	0	2			أ	ح	ب	ط	ك	و	
0	0	1	1	3			ط	ح	ز	ط	ك	و	
0	1	0	0	4			ط	ط	ش	ط	ل	ي	
0	1	0	1	5			ن	ه	ت	ط	ل	ي	
0	1	1	0	6			ب	ب	ب	ف	ث	ي	
0	1	1	1	7			ه	ط	ب	ط	ل	ي	
1	0	0	0	8			ن	ط	ب	ي	ر	ع	
1	0	0	1	9			ب	ي	ش	ط	ث	⊞	
1	0	1	0	A			ث	ط	ب	ي	ب	⊞	
1	0	1	1	B			ب	ط	ش	ك	ث	⊞	
1	1	0	0	C			ع	ط	ب	ك	ب	⊞	
1	1	0	1	D			خ	ط	ظ	ل	ه	⊞	
1	1	1	0	E			ح	ر	ع	ك	ه	⊞	
1	1	1	1	F			ح	ر	خ	ي	و		

⊞ Reserved for future standardization

5 Legend

Table 2 gives the code, graphic and name of each character and comments on usage when needed.

Table 2

Code	Graphic	Name	Comments
21	أأ	ARABIC LETTER DOUBLE ALEF WITH HAMZAH ABOVE	Sindhi ampersand
22	آ	ARABIC LETTER ALEF WITH WAVY HAMZAH ABOVE	Used in Baluchi
23	اٲ	ARABIC LETTER ALIF WITH WAVY HAMZAH BELOW	Used in Baluchi
24	ٹ	ARABIC LETTER TTEH	Used in Urdu
25	ٲ	ARABIC LETTER TTEHEH	Used in Sindhi
26	بٲ	ARABIC LETTER BEEH	Used in Sindhi
27	ٲٲ	ARABIC LETTER TEH WITH RING	Used in Pushto
28	ٲٲٲ	ARABIC LETTER TEH WITH THREE DOTS ABOVE DOWNWARD	Used in Sindhi
29	پ	ARABIC LETTER PEH	Used in Farsi, etc.
2A	ٲٲٲ	ARABIC LETTER TEHEH	Used in Sindhi
2B	ٲٲٲ	ARABIC LETTER BEHEH	Used in Sindhi
2C	حٲ	ARABIC LETTER HAH WITH HAMZAH ABOVE	Used in Pushto
2D	حٲٲ	ARABIC LETTER HAH WITH TWO DOTS VERTICAL ABOVE	Used in Pushto
2E	ٲٲٲ	ARABIC LETTER NYEH	Used in Sindhi
2F	ٲٲٲ	ARABIC LETTER DYEH	Used in Sindhi
30	حٲٲٲ	ARABIC LETTER HAH WITH THREE DOTS ABOVE	Used in Pushto
31	ٲٲٲٲ	ARABIC LETTER TCHEH	Used in Farsi, etc.
32	ٲٲٲٲ	ARABIC LETTER TCHEH WITH DOT ABOVE	Used in Kurdish
33	ٲٲٲٲٲ	ARABIC LETTER TCHEHEH	Used in Sindhi
34	ڈ	ARABIC LETTER DDAL	Used in Urdu
35	دٲ	ARABIC LETTER DAL WITH RING	Used in Pushto
36	دٲٲ	ARABIC LETTER DAL WITH DOT BELOW	Used in Sindhi
37	دٲٲٲ	ARABIC LETTER DAL WITH DOT BELOW AND TAH ABOVE	Used in Lahnda
38	دٲٲٲ	ARABIC LETTER DAHAL	Used in Sindhi
39	دٲٲٲٲ	ARABIC LETTER DDAHAL	Used in Sindhi
3A	ذ	ARABIC LETTER DUL	Used in Sindhi
3B	ذٲٲٲ	ARABIC LETTER DAL WITH THREE DOTS ABOVE DOWNWARD	Used in Sindhi
3C	ذٲٲٲٲ	ARABIC LETTER DAL WITH FOUR DOTS ABOVE	Used in Urdu
3D	ڑ	ARABIC LETTER RREH	Used in Urdu
3E	رٲ	ARABIC LETTER REH WITH CARON ABOVE	Used in Kurdish
3F	رٲٲٲ	ARABIC LETTER REH WITH RING	Used in Pushto

Table 2 (continued)

Code	Graphic	Name	Comments
40	ر	ARABIC LETTER REH WITH DOT BELOW	Used in Kurdish
41	ر̣	ARABIC LETTER REH WITH CARON BELOW	Used in Kurdish
42	ر̇	ARABIC LETTER REH WITH DOT ABOVE AND DOT BELOW	Used in Pushto
43	ز	ARABIC LETTER REH WITH TWO DOTS ABOVE	Used in Dargwa
44	ژ	ARABIC LETTER JEH	Used in Farsi, etc.
45	ز̇	ARABIC LETTER REH WITH FOUR DOTS ABOVE	Used in Sindhi
46	ښ	ARABIC LETTER SEEN WITH DOT ABOVE AND DOT BELOW	Used in Pushto
47	پس	ARABIC LETTER SEEN WITH THREE DOTS BELOW	Used in Uighur
48	پس̇	ARABIC LETTER SEEN WITH THREE DOTS ABOVE AND THREE DOTS BELOW	Used in Berber
49	ش	ARABIC LETTER SHEEN WITH DOT BELOW	Used in Moplah
4A	س̇	ARABIC LETTER SAD WITH TWO DOTS BELOW	Used in Turkish
4B	ش̇	ARABIC LETTER SAD WITH THREE DOTS ABOVE	Used in Berber
4C	ص̇	ARABIC LETTER DAD WITH DOT BELOW	Used in Moplah
4D	ظ	ARABIC LETTER TAH WITH THREE DOTS ABOVE	Used in Hausa
4E	ع̇	ARABIC LETTER AIN WITH THREE DOTS ABOVE	Used in Malay
4F	غ	ARABIC LETTER GHAIN WITH DOT BELOW	Used in Moplah
50	ف	ARABIC LETTER DOTLESS FEH	Used in Adighe
51	ف̣	ARABIC LETTER FEH WITH DOT MOVED BELOW	Used in Berber
52	ف̇	ARABIC LETTER FEH WITH DOT BELOW	Used in Turkish
53	ف̈	ARABIC LETTER VEH	Used in various languages
54	ف̇̈	ARABIC LETTER DOTLESS FEH WITH THREE DOTS BELOW	Used in various languages
55	ف̈̈	ARABIC LETTER PEHEH	Used in Sindhi
56	ق̇	ARABIC LETTER QAF WITH DOT ABOVE	Used in Berber
57	ق̇̈	ARABIC LETTER QAF WITH THREE DOTS ABOVE	Used in Berber
58	ق	ARABIC LETTER KEHEH	Used in Pushto
59	ڪ	ARABIC LETTER SWASH CAF	Used in Sindhi
5A	ڪ̇	ARABIC LETTER KAF WITH RING	Used in Pushto
5B	ڪ̈	ARABIC LETTER CAF WITH DOT ABOVE	Used in Malay
5C	ڪ̇̈	ARABIC LETTER NG	Used in Malay
5D	ڪ̇̈̈	ARABIC LETTER CAF WITH THREE DOTS BELOW	Used in Berber
5E	گ	ARABIC LETTER GAF	Used in Farsi, etc.
5F	گ̇	ARABIC LETTER GAF WITH RING	Used in Lahnda

Table 2 (concluded)

Code	Graphic	Name	Comments
60	ڱ	ARABIC LETTER NGOEH	Used in Sindhi
61	گي	ARABIC LETTER GAF WITH TWO DOTS BELOW	Used in Sindhi
62	گپ	ARABIC LETTER GUEH	Used in Sindhi
63	گا	ARABIC LETTER GAF WITH THREE DOTS ABOVE	Used in Sindhi
64	ل̇	ARABIC LETTER LAM WITH CARON ABOVE	Used in Kurdish
65	ل̈	ARABIC LETTER LAM WITH DOT ABOVE	Used in Kurdish
66	ل̉	ARABIC LETTER LAM WITH THREE DOTS ABOVE	Used in Kurdish
67	ل̋	ARABIC LETTER LAM WITH THREE DOTS BELOW	Used in Avaric
68	ن	ARABIC LETTER NOON GHUNNA	Used in Urdu
69	ن̇	ARABIC LETTER RNOON	Used in Sindhi
6A	ن̈	ARABIC LETTER NOON WITH RING	Used in Pushto
6B	ن̉	ARABIC LETTER NOON WITH THREE DOTS	Used in Malay
6C	ن̋	ARABIC LETTER NOON WITH DOT BELOW	Used in Moplah
6D	هھ	ARABIC LETTER HEH DOACHASHMEE	Used in Urdu
6E	ه̈	ARABIC LETTER HAMZAH ON HA	Used in Farsi
6F	و̈	ARABIC LETTER WAW WITH RING	Used in Kashmiri
70	و̇	ARABIC LETTER KIRGHIZ OE	Used in Kirghiz
71	ۆ	ARABIC LETTER OE	Used in Kurdish
72	ۆ̇	ARABIC LETTER WAW WITH TWO DOTS	Used in Kurdish
73	ۆ̈	ARABIC LETTER KIRGHIZ YU	Used in Uighur
74	ه̇	ARABIC LETTER YEH WITH TAIL	Used in Sindhi
75	ه̈	ARABIC LETTER YA WITH CARON ABOVE	Used in Kurdish
76	ه̉	ARABIC LETTER E	Used in Pushto
77	ه̋	ARABIC LETTER YEH BARREE	Used in Urdu
78	ع̈	ARABIC LETTER PERIOD	Used in Urdu
79		(This position is not used)	
7A		(This position is not used)	
7B		(This position is not used)	
7C		(This position is not used)	
7D	ۀ	ARABIC LETTER SHORT E	Used in Urdu
7E	ؤ	ARABIC LETTER SHORT U	Used in Urdu

6 Explanatory notes

6.1 The 7-bit code table (table 1) 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 F.

The code table positions are identified by notations of the form xy , 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 xy , is the same as that of the corresponding bit combination.

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

6.2 Certain vowels, generally short vowels, are represented in the Arabic script by special vowel marks. These vowel marks are always used in conjunction with other graphic characters.

ISO 9036 includes the most commonly used vowel marks. This International Standard includes two additional marks, in character positions 7D and 7E, for short vowels used in Urdu. The vowel mark allocated to position 7E is also occasionally used to differentiate certain consonants.

6.3 The characters in positions 7D and 7E are designated as non-spacing graphic characters, that is, characters whose use is not followed by the forward movement of the output device. In a character string, these non-spacing characters are input before the characters they modify.

6.4 The rendering of graphic characters is intended solely to identify the additional letters of the Arabic alphabet uniquely. The graphics used do not necessarily represent the most desirable calligraphic forms.

6.5 The names of characters (but not codes) have been made to correspond as much as possible to those assigned in ISO/IEC 10646-1.

Annex A
(informative)

Basic Arabic character set table from ISO 9036

				b ₇	0	0	0	0	1	1	1	1
				b ₆	0	0	1	1	0	0	1	1
				b ₅	0	1	0	1	0	1	0	1
b ₄	b ₃	b ₂	b ₁		0	1	2	3	4	5	6	7
0	0	0	0	0	NUL	DLE	SP	0	@	ذ	—	ء
0	0	0	1	1	SOH	DC1	!	1	ء	ر	ف	آ
0	0	1	0	2	STX	DC2	"	2	آ	ز	ق	٠
0	0	1	1	3	ETX	DC3	#	3	أ	س	ك	⊠
0	1	0	0	4	EOT	DC4	␣	4	ؤ	ش	ل	⊠
0	1	0	1	5	ENQ	NAK	%	5	!	ص	م	⊠
0	1	1	0	6	ACK	SYN	&	6	ئ	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	A	LF	SUB	*	:	ت	غ	ي	⊠
1	0	1	1	B	VT	ESC	+	؛	ث]	=	}
1	1	0	0	C	FF	IS4	,	>	ج	\	”	
1	1	0	1	D	CR	IS3	-	=	ح	[=	{
1	1	1	0	E	SO	IS2	.	<	خ	^	’	~
1	1	1	1	F	SI	IS1	/	?	د	_	’	DEL

⊠ Reserved for future standardization

Annex B

(informative)

Bibliography

- [1] ISO 962:1974, *Information processing — Implementation of the 7-bit coded character set and its 7-bit and 8-bit extensions on 9-track 12,7 mm (0.5 in) magnetic tape.*
- [2] ISO 1155:1978, *Information processing — Use of longitudinal parity to detect errors in information messages.*
- [3] ISO 1177:1985, *Information processing — Character structure for start/stop and synchronous character oriented transmission.*
- [4] ISO 1745:1975, *Information processing — Basic mode control procedures for data communication systems.*
- [5] ISO/IEC 10646-1:1993, *Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane.*

ICS 35.040

Descriptors: documentation, bibliographies, data processing, information interchange, graphic characters, Arabic characters, character sets, coded character sets, extensions.

Price based on 9 pages
