International Standard ISO 5426

Extension of the Latin alphabet coded character set for bibliographic information interchange

Extension du jeu de caractères latins codés employé pour l'échange d'information bibliographique

Second edition — 1983-11-01

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 5426 was developed by Technical Committee ISO/TC 46, *Documentation*.

This second edition was submitted directly to the ISO Council, in accordance with clause 6.11.2 of part 1 of the Directives for the technical work of ISO. It cancels and replaces the first edition (i.e. ISO 5426-1980), which had been approved by the member bodies of the following countries :

Austria	Hungary	Poland
Belgium	India	Romania
Brazil	Ireland	South Africa, Rep. of
Canada	Israel	Spain
Czechoslovakia	Italy	Sweden
Denmark	Japan	Switzerland
Egypt, Arab Rep. of	Korea, Rep. of	United Kingdom
Finland	Mexico	USA
France	Netherlands	USSR
Germany, F. R.	Norway	Yugoslavia

No member body had expressed disapproval of the document.

Extension of the Latin alphabet coded character set for bibliographic information interchange

1 Scope and field of application

1.1 This International Standard contains a set of 76 graphic characters with coded representations. It consists of a code table and a legend showing each graphic, its use, and its name. Explanatory notes are also included. This 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 646 (ISO escape sequence ESC 2/8 4/0), constitute a character set for the international interchange of bibliographic citations, including their annotations, in the Latin alphabet.

1.3 This character set is intended to handle information in the following languages, among others :

African languages (those using Latin alphabet)[1]	Interlingua
	Italian
	Ladino
Albanian	Latin
Anglo-Saxon	Latvian
Catalan	Lithuanian
Croatian	Norwegian
Czech	Polish
Danish	Portuguese
Dutch	Romanian
English	Samoan
Esperanto	Slovak
Estonian	Slovenian
Finnish	Spanish
French	Swedish
Gaelic	Tagalog
German	Turkish (modern)
Hawaiian	Vietnamese
Hungarian	Welsh
Icelandic (modern)	Wendish
Indonesian/Malayan	

It is also intended to handle transliterated or romanized forms of the following languages, among others :

Arabic	Japanese	Sanskrit
Armenian	Kannada	Serbian
Byelorussian	Korean	Sindhi
Bengali	Macedonian	Sinhalese
Bulgarian	Marathi	Tamil
Burmese	Oriya	Telugu
Chinese	Panjabi	Thai

Greek	Persian	Tibetan
Gujarati	Prakrit	Ukrainian
Hebrew	Pushto	Urdu
Hindi	Russian	

2 References

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

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

ISO 2375, *Data processing — Procedure for registration of escape sequences.*

ISO 6438, *Documentation — African coded character set for bibliographic information interchange.*[2]

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 ISO publications. (See clause 7.)

3.2 The implementation of this International Standard is in accordance with the provisions of ISO 2022 and ISO 2375 and is identified by the following escape sequence :

```
G0 : ESC 2/8  5/0
G1 : ESC 2/9  5/0
G2 : ESC 2/10 5/0
G3 : ESC 2/11 5/0
```

3.3 The unassigned positions in the code table are not to be utilized in the international exchange of bibliographic information.

3.4 When this character set is used in combination with ISO 646 [International Reference Version (IRV)] as stated in 1.2 the following character positions of ISO 646 shall not be used (the corresponding positions in this International Standard are noted within parentheses) :

5/15 (5/8); 6/0 (4/1); 7/14 (4/5)

Also, characters in positions 2/2, 2/7, 2/12 and 5/14 in ISO 646 may not be used as diacritics.

---

1) For special characters which occur in African languages only, see ISO 6438.

2) At present at the stage of draft.

4 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
b4 b3 b2 b1			**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0 0 0 0	0					'	ˀ	⸴		
0 0 0 1	1				¡	'	`	¿	Æ	æ
0 0 1 0	2				,,	,	´	⌡	Đ	đ
0 0 1 1	3				£		^	⌣		ð
0 1 0 0	4				$			°		
0 1 0 1	5				¥			˘		ł
0 1 1 0	6				†	‡	˘	.	IJ	ij
0 1 1 1	7				§	·	˙	..		
1 0 0 0	8				'	''	¨	_	Ł	ł
1 0 0 1	9				'	'	¨	=	Ø	ø
1 0 1 0	10				"	"	°	ˏ	Œ	œ
1 0 1 1	11				«	»	'	^		ß
1 1 0 0	12				♭	#	'		Þ	þ
1 1 0 1	13				©	´	"	⌐		
1 1 1 0	14				℗	"	'	⌐		
1 1 1 1	15				®	¿	ˇ	‿		

5 Legend

5.1 Symbols

Position in code table	Graphic	Example of use	Name	Comments
2/0			Space	
2/1	¡	¡Pare	Inverted exclamation mark	Used in Spanish
2/2	,,	,,Unter der . . .	Double open quote	Prints low on line. Used in German, etc.
2/3	£	£5.00	Pound sign	
2/4	$	$2.00	Dollar sign	
2/5	¥	¥300	Yen sign	
2/6	†	word†	Single dagger	Reference mark
2/7	§	§ 5	Section or paragraph mark	
2/8	'	50' (minutes of angle; feet; x' = first derivative)	Prime	
2/9	'	'Four score . . .	Single open quote	Prints high on line
2/10	''	''Hurry!	Double open quote	Prints high on line
2/11	«	«C'est la vie.	Angle quote	Either open or close quote according to national usage
2/12	♭	B♭	Musical flat	
2/13	©	©1974	Copyright	
2/14	℗	℗ 1973 Doe Records Inc.	Sound recording copyright statement	
2/15	®	XYZ®	Registered trade mark	
3/0	'	T'ai-pei	Ayn	Not to be used as a single open quote.
3/1	'	Un'yushō	Alif/Hamzah	Not to be used as a single close quote.
3/2	,	,Über. . .	Single open quote	Prints low on line. Used in German, etc.
3/6	‡	word ‡	Double dagger	Reference mark
3/7	·	novel·la	Middle dot	Used in Catalan
3/8	''	8'' (seconds of angle; inches; x'' = second derivative)	Double prime	
3/9	'	Four score. . .'	Single close quote	Prints high on line
3/10	''	Hurry!''	Double close quote	Prints high on line
3/11	»	C'est la vie.»	Angle quote	Either open or close quote according to national usage
3/12	♯	C ♯	Musical sharp	
3/13	'	Fakul'tet	Mjagkij znak	Used in transliterated Cyrillic alphabet languages
3/14	''	Ob''iavlenie	Tverdyj znak	Used in transliterated Cyrillic alphabet languages
3/15	¿	¿Qué	Inverted question mark	Used in Spanish

3

5.2 Diacritics

Position in code table	Graphic	Example of use	Name	Comments
4/0	ʔ	cửi	Low rising tone mark	Used in Vietnamese
4/1	`	règle	Grave accent	Used in many European languages, etc.
4/2	´	está	Acute accent	Used in many European languages, etc.
4/3	^	même	Circumflex accent	Used in many European languages, etc.
4/4	~	niño	Tilde	Used in Portuguese, Spanish, etc.
4/5	‾	gājējs	Macron	Used in Latvian, Lithuanian, etc.
4/6	˘	altă	Breve	Used in Romanian, Turkish, etc.
4/7	·	żaba	Dot above	Used in Lithuanian, Polish, transliterated Arabic, etc.
4/8	¨	Zaïre	Trema, Diaeresis	These diacritics may have different filing values. 4/8 is used in French, etc. 4/9 is used in German, etc.
4/9	¨	är, öffnen, öppna, über	Umlaut	
4/10	°	hår	Circle above	Used in Scandinavian languages, etc.
4/11	ʼ	rozdeľovac	High comma off centre	Used in Czech, Slovak, etc.
4/12	ʻ	ģeotermika	High inverted comma centred	Used in Latvian
4/13	˝	időszaki	Double acute accent	Used in Hungarian
4/14	ʾ	xưa	Horn	Used in Vietnamese and transliterated Thai
4/15	ˇ	vždy	Háček	Used in Czech, Latvian, etc.
5/0	̧	ça	Cedilla	Used in French, etc.
5/1	̢	kh̥ng	Rude	Used in transliterated Thai
5/2	̦	darziņa	Hook to left	Used in Latvian, Romanian
5/3	̨	vietą	Hook to right or ogonek	Used in Lithuanian, Polish, etc.
5/4	̥	Maharṣicaritāmṛtam	Circle below	Used in transliterated Bengali, Hindi, etc.
5/5	̯	ḫumantuš	Half circle below	Used in transliterated Semitic languages, Sanskrit
5/6	̣	teḍa	Dot below	Used in transliterated Bengali, Hindi, etc.
5/7	̤	Khuṭbah	Double dot below	Used in transliterated Urdu
5/8	̲	samar	Underline	Used in transliterated Bengali, Hindi, Urdu, etc.

4

Position in code table	Graphic	Example of use	Name	Comments
5/9	=	G̱hulam	Double underline	Used in transliterated Hindi, Sindhi
5/10	ˌ	lati ọwọ	Vertical bar	Used in African languages
5/11	ˆ	nję̂n	Circumflex below	Used in African languages
5/13	⌐	Akademiîa	Left half of ligature sign and of double tilde	Both ligature and double tilde are divided into two parts, the first parts of each being identical. Used in trans-literations
5/14	˥	Akademiîa	Right half of ligature sign	
5/15	⌣	ñgalan	Right half of double tilde	Used in Tagalog

5.3 Special alphabetic characters

Position in code table	Graphic	Example of use	Name	Comments
6/1	Æ	Ætna	Ligature AE — Capital letter	Used in Danish, etc.
6/2	Đ	Đuro	Slash D — Capital letter	Used in Croatian, Icelandic, etc.
6/6	IJ	IJsselmeer	Digraph IJ — Capital letter	Used in Dutch
6/8	Ł	Łódź	Slash L — Capital letter	Used in Polish, Wendish
6/9	Ø	Øst	Slash O — Capital letter	Used in Danish, Norwegian, etc.
6/10	Œ	Œuvre	Ligature OE — Capital letter	Used in French, etc.
6/12	Þ	Þann	Thorn — Capital letter	Used in Icelandic, etc.
7/1	æ	skæg	Ligature ae — Small letter	Used in Danish, etc.
7/2	đ	đavola	Slash d — Small letter	Used in Croatian, transliterated Serbian, etc.
7/3	ð	verður	eth — Small letter	Used in Anglo-Saxon, Ice-landic
7/5	ı	masalı	Dotless i — Small letter	Used in Turkish
7/6	ij	Rijks	Ligature ij — Small letter	Used in Dutch
7/8	ł	rozbił	Slash l — Small letter	Used in Polish, Wendish
7/9	ø	høj	Slash o — Small letter	Used in Danish, Norwegian, etc.
7/10	œ	œuvre	Ligature oe — Small letter	Used in French, etc.
7/11	ß	Schloß	Ligature sz — Small letter	Used in German
7/12	þ	þann	Thorn — Small letter	Used in Icelandic, etc.

6 Explanatory notes

6.1 Three kinds of characters are included in the set :

a) Symbols (for example punctuation, pound sign, dollar mark, etc.). These characters appear in columns 2 and 3.

b) Diacritics (for example tilde, acute accent, cedilla, etc.). These characters are always used in conjunction with other characters and appear in columns 4 and 5. Diacritics always precede the character which they modify. Multiple diacritics are to be entered in the order in which they appear reading left to right or top to bottom. The BACKSPACE character (0/8 of ISO 646) is not used to identify diacritics.

c) Special alphabetic characters (for example Icelandic thorn). These characters appear in columns 6 and 7.

6.2 Characters composed of two (or more) components, of which one is a qualifying mark, are encoded as follows :

a) Characters in which a qualifying mark is superimposed upon the basic character are encoded as a single character (for example ø, Đ).

b) Characters in which the qualifying mark is separated from, or only touches, the basic character are encoded as two characters (for example Ç = ‚ and C; Å = ° and A).

6.3 Characters not based on alphabetic characters in ISO 646 are encoded as single characters (for example #).

6.4 Wherever possible, characters appearing as both capital and small letters are given codes differing in the value of only one bit position. This principle applies only where there is a one-to-one relationship and is not extended to the dotless i, the Icelandic eth, and the German ß.

6.5 In general, separate printed shapes are considered separate graphic characters. Where a concept can be expressed by several different characters or compressed into a single character, the decision has been to designate as many characters as needed (for example several types of quotation marks are included). However, different font designs are not considered separate characters. Such fonts will be designated by control characters.

6.6 The diacritic characters underline, overline, acute accent, grave accent, circumflex accent, diaeresis, and cedilla are not to be used in the ISO 646 (IRV) set and are therefore repeated in the extended set (see 2.4).

7 Bibliography

[1] ISO 962, *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 1113, *Information processing — Representation of the 7-bit coded character set on punched tape.*

[3] ISO 1155, *Information processing — Use of longitudinal parity to detect errors in information messages.*

[4] ISO 1177, *Information processing — Character structure for start/stop and synchronous transmission.*

[5] ISO 1745, *Information processing — Basic mode control procedures for data communication systems.*

[6] ISO 2375, *Data processing — Procedure for registration of escape sequences.*

[7] ISO 6586, *Data processing — Implementation of the ISO 7-bit and 8-bit coded character sets on punched cards.*