

INTERNATIONAL STANDARD

ISO
10754

First edition
1996-12-15

Information and documentation — Extension of the Cyrillic alphabet coded character set for non-Slavic languages for bibliographic information interchange

*Information et documentation — Extension du jeu de caractères codés de
l'alphabet cyrillique pour les langues non-slaves pour les échanges
d'informations bibliographiques*



Reference number
ISO 10754:1996(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 10754 was prepared by Technical Committee ISO/TC 46, *Information and documentation*, Subcommittee SC 4, *Computer applications in information and documentation*.

Annexes A, B and C 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 Cyrillic alphabet coded character set for non-Slavic languages for bibliographic information interchange

1 Scope

1.1 This International Standard specifies a set of 93 graphic characters with their coded representations. It consists of a code table and a legend showing each graphic, its use and its name. 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 basic Cyrillic set, registered as number 37 in the ISO international register, constitute a character set for the international interchange of bibliographic citations, including their annotations, in the non-Slavic Cyrillic alphabets for the languages specified in 1.3.

1.3 This character set is intended to handle information in the following language groups:

Abazian	Kabardian	Mordvin
Abkhasian	Kalmyk	Nenets
Adyghe	Karachay	Nivkh
Aisor	Kara-Kalpak	Nogai
Altaic	Karelian	Ossetic
Avar	Kazakh	Romany
Azerbaijani	Khakass	Sami
Balkar	Khanty	Selkup
Bashkir	Kirghiz	Shor
Buryat	Komi	Tabasaran
Chechen	Koryak	Tajik
Chukchi	Kumyk	Tat
Chuvash	Kurdish	Tatar
Dargwa	Lak	Turkmen
Dungan	Lezghian	Tuvinian
Eskimo	Lithuanian	Udekhe
Even	Mansi	Udmurt
Evenki	Mari	Uighur
Gagauzi	Moldavian	Uzbek
Ingush	Mongolian	Yakut

1.4 This coded character set contains characters used since the Russian Revolution (1917). Some letters which appear to be unrepresented in the character table are actually graphic variants. Obsolete letters, those used for only a brief period in the late 19th century, have been excluded from this International Standard. This applies chiefly to early letters used in Chechen, Chuvash, Dargwa, Lak and Lezghian. Letters from their 20th century alphabets are included.

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 646:1991, *Information technology — ISO 7-bit coded character set for information interchange*.

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

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

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 C).

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 tables shall not be utilized in the international interchange of bibliographic information.

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

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 Cyrillic characters of non-Slavic languages

Table 1 is the code table for extended Cyrillic characters of non-Slavic languages.

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					⌘	і	ї	ѣ	Ѧ
0	0	0	1	1				ђ	ѓ	ј	ђ	ѡ	Ѣ
0	0	1	0	2				Ѣ	ѣ	ѡ	Ѣ	ѣ	Ѣ
0	0	1	1	3				ѣ	ѣ	ѡ	Ѣ	ѣ	Ѣ
0	1	0	0	4				ѣ	ѣ	ѡ	Ѣ	ѣ	Ѣ
0	1	0	1	5				ѣ	ѣ	ѡ	Ѣ	ѣ	Ѣ
0	1	1	0	6				ѣ	ѣ	ѡ	Ѣ	ѣ	Ѣ
0	1	1	1	7				ѣ	ѣ	ѡ	Ѣ	ѣ	Ѣ
1	0	0	0	8				ѣ	ѣ	ѡ	Ѣ	ѣ	Ѣ
1	0	0	1	9				ѣ	ѣ	ѡ	Ѣ	ѣ	Ѣ
1	0	1	0	A				ѣ	ѣ	ѡ	Ѣ	ѣ	Ѣ
1	0	1	1	B				ѣ	ѣ	ѡ	Ѣ	ѣ	Ѣ
1	1	0	0	C				ѣ	ѣ	ѡ	Ѣ	ѣ	Ѣ
1	1	0	1	D				ѣ	ѣ	ѡ	Ѣ	ѣ	Ѣ
1	1	1	0	E				ѣ	ѣ	ѡ	Ѣ	ѣ	Ѣ
1	1	1	1	F				ѣ	ѣ	ѡ	Ѣ	ѣ	Ѣ

⌘ Reserved for future standardization

5 Legend

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

Table 2

Code	Graphic	Name	Comments
21	´	COMBINING ACUTE	
22	¨	COMBINING DIAERESIS (Dialytika)	
23	˚	COMBINING OGONEK	
24	¸	COMBINING RIGHT DESCENDER	
25	˘	COMBINING BREVE	
26	¸	COMBINING CEDILLA	
27	̀	COMBINING GRAVE	
28	Ѧ	CYRILLIC SMALL LETTER A IE	
29	Ѣ	CYRILLIC SMALL LETTER GHE WITH STROKE	
2A	Ѥ	CYRILLIC SMALL LETTER GHE WITH MIDDLE HOOK	
2B	Ѧ	CYRILLIC SMALL LETTER KOMI DE	
2C	ѧ	CYRILLIC SMALL LETTER KOMI DJE	
2D	Ӏ	CYRILLIC SMALL LETTER ABKHASIAN DZE	
2E	ӑ	CYRILLIC SMALL LETTER KOMI DZE	
2F	ӕ	CYRILLIC SMALL LETTER KOMI ZJE	
30		(This position shall not be used)	
31	ˆ	COMBINING DOUBLE ACUTE	
32	˜	COMBINING MACRON	
33	˚	COMBINING LEFT OGONEK	
34	¸	COMBINING LEFT DESCENDER	
35	ˇ	COMBINING CARON	
36	ˆ	COMBINING RING ABOVE	
37	ˆ	COMBINING HIGH COMMA	
38	Ѧ	CYRILLIC CAPITAL LETTER A IE	
39	Ѣ	CYRILLIC CAPITAL LETTER GHE WITH STROKE	
3A	Ѥ	CYRILLIC CAPITAL LETTER GHE WITH MIDDLE HOOK	
3B	Ѧ	CYRILLIC CAPITAL LETTER KOMI DE	
3C	ѧ	CYRILLIC CAPITAL LETTER KOMI DJE	
3D	Ӏ	CYRILLIC CAPITAL LETTER ABKHASIAN DZE	
3E	ӑ	CYRILLIC CAPITAL LETTER KOMI DZE	
3F	ӕ	CYRILLIC CAPITAL LETTER KOMI ZJE	

Table 2 (continued)

Code	Graphic	Name	Comments
40	і	CYRILLIC SMALL LETTER YAKUT I WITH STROKE	
41	ј	CYRILLIC SMALL LETTER JE WITH STROKE	
42	к	CYRILLIC SMALL LETTER KA WITH VERTICAL STROKE	
43	к	CYRILLIC SMALL LETTER BASHKIR KA	
44	к	CYRILLIC SMALL LETTER KA WITH STROKE	
45	к	CYRILLIC SMALL LETTER CHECHEN KA	Also in Dargwa, Lak, Lezghian
46	қ	CYRILLIC SMALL LETTER KURDISH QA	
47	л	CYRILLIC SMALL LETTER AISOR EL	Also in Yakut
48	л	CYRILLIC SMALL LETTER KOMI ELJ	
49	л	CYRILLIC SMALL LETTER EL WITH MIDDLE HOOK	
4A	л	CYRILLIC SMALL LETTER MORDVIN EL KA	
4B	н	CYRILLIC SMALL LETTER ALTAIC NG	Also Mordvin and Yakut
4C	н	CYRILLIC SMALL LETTER CHUVASH NG	
4D	н	CYRILLIC SMALL LETTER KOMI NG	
4E	н	CYRILLIC SMALL LETTER EN WITH MIDDLE HOOK	
4F	о	CYRILLIC SMALL LETTER O WITH STROKE	
50	І	CYRILLIC CAPITAL LETTER YAKUT I WITH STROKE	
51	Ј	CYRILLIC CAPITAL LETTER JE WITH STROKE	
52	К	CYRILLIC CAPITAL LETTER KA WITH VERTICAL STROKE	
53	К	CYRILLIC CAPITAL LETTER BASHKIR KA	
54	К	CYRILLIC CAPITAL LETTER KA WITH STROKE	
55	К	CYRILLIC CAPITAL LETTER CHECHEN KA	Also in Dargwa, Lak, Lezghian
56	Q (Q)	CYRILLIC CAPITAL LETTER KURDISH QA	"Q" is alternate rendering
57	Л	CYRILLIC CAPITAL LETTER AISOR EL	Also used in Yakut
58	Л	CYRILLIC CAPITAL LETTER KOMI ELJ	
59	Л	CYRILLIC CAPITAL LETTER EL WITH MIDDLE HOOK	
5A	Л	CYRILLIC CAPITAL LETTER MORDVIN EL KA	
5B	Н	CYRILLIC CAPITAL LETTER ALTAIC NG	Also used in Mordvin and Yakut
5C	Н	CYRILLIC CAPITAL LETTER CHUVASH NG	
5D	Н	CYRILLIC CAPITAL LETTER KOMI NG	
5E	Н	CYRILLIC CAPITAL LETTER EN WITH MIDDLE HOOK	
5F	Ө	CYRILLIC CAPITAL LETTER O WITH STROKE	

Table 2 (concluded)

Code	Graphic	Name	Comments
60	Ɑ	CYRILLIC SMALL LETTER ABKHASIAN HA	
61	Ɱ	CYRILLIC SMALL LETTER SELKUP O IE	
62	Ɐ	CYRILLIC SMALL LETTER ABKHASIAN PHE	
63	Ɒ	CYRILLIC SMALL LETTER ER KA	
64	ⱱ	CYRILLIC SMALL LETTER KOMI ESJ	
65	Ⱳ	CYRILLIC SMALL LETTER KOMI TJE	
66	ⱳ	CYRILLIC SMALL LETTER STRAIGHT U	
67	ⱴ	CYRILLIC SMALL LETTER STRAIGHT U WITH STROKE	
68	Ⱶ	CYRILLIC SMALL LETTER KURDISH WE	Also used in Dargwa, Lezghian
69	ⱶ	CYRILLIC SMALL LETTER ABKHASIAN THE	
6A	ⱷ	CYRILLIC SMALL LETTER CHE WITH VERTICAL STROKE	
6B	ⱸ	CYRILLIC SMALL LETTER HE	
6C	ⱹ	CYRILLIC SMALL LETTER ABKHASIAN CHE	
6D	ⱺ	CYRILLIC SMALL LETTER SHWA	
6E	ⱻ	CYRILLIC SMALL LETTER YA IE	
6F	ⱼ	CYRILLIC ASPIRATION OR GUTTURAL SIGN	Used in Caucasian languages
70	ⱼ	CYRILLIC CAPITAL LETTER ABKHASIAN HA	
71	ⱽ	CYRILLIC CAPITAL LETTER SELKUP O IE	
72	Ȿ	CYRILLIC CAPITAL LETTER ABKHASIAN PHE	
73	Ɀ	CYRILLIC CAPITAL LETTER ER KA	
74	Ȿ	CYRILLIC CAPITAL LETTER KOMI ESJ	
75	Ɀ	CYRILLIC CAPITAL LETTER KOMI TJE	
76	Ɀ	CYRILLIC CAPITAL LETTER STRAIGHT U	
77	Ɀ	CYRILLIC CAPITAL LETTER STRAIGHT U WITH STROKE	
78	Ɀ	CYRILLIC CAPITAL LETTER KURDISH WE	Also used in Dargwa, Lezghian
79	Ɀ	CYRILLIC CAPITAL LETTER ABKHASIAN THE	
7A	Ɀ	CYRILLIC CAPITAL LETTER CHE WITH VERTICAL STROKE	
7B	Ɀ	CYRILLIC CAPITAL LETTER HE	
7C	Ɀ	CYRILLIC CAPITAL LETTER ABKHASIAN CHE	
7D	Ɀ	CYRILLIC CAPITAL LETTER SHWA	
7E	Ɀ	CYRILLIC CAPITAL LETTER YA IE	

6 Explanatory notes

6.1 Punctuation marks and numerals in European style used in the non-Slavic languages covered by this International Standard are available in the basic Cyrillic set (Registration No. 37 in the international register with which this set is designed for use).

6.2 In many of the non-Slavic languages, diacritical marks are combined with Cyrillic script letters to create distinctive modified letters. These marks are usually placed above or below a letter. The most common marks include the diaeresis (Ѣ), the acute (Ё), right descender (Ж), left descender (Ў), ogonek (Ъ) and left ogonek (Ы). The use of such marks is widespread and thus, several non-spacing combining marks are defined in this International Standard.

In some texts, a *high comma* is used above letters instead of an *acute* mark (Ё). Most modern Cyrillic script reference sources tend to use the acute, however. Both combining marks are defined in this International Standard. In older texts, an apostrophe is occasionally used to represent modified letters. When this character is needed, the apostrophe provided in the basic Cyrillic set (Registration No. 37 in the international register) should be used.

The non-Slavic languages make liberal and sometimes inconsistent use of the *right descender*, *left descender*, *ogonek* and *left ogonek*, in combination with many letters, especially consonants, to show palatalization, aspiration, etc. The *cedilla* is also combined with some consonants to represent certain sounds. These combining marks have been defined in this International Standard to permit the encoding of the large number of combinations that have been identified.

Characters with large middle hooks or tails are defined as separate characters. Sources identify these marks as either: "hvostik" (tail), "sedil'" (cedilla), or "krjuk" (hook).

6.3 The guttural or aspiration sign (*pridykhatel'nyj znak*; Ё) must not be confused with the Latin script capital "I". This sign is used in many Caucasian languages. It always follows a consonant and has the same form regardless of the case of the other letters in a word (e.g. *rlanah*). Although technically a sign (like a percent "%" sign), this character is given as the last letter of most Cyrillic-based alphabets. The notion of capitalization is not applied to this sign, thus, it is assigned only one code in this International Standard.

6.4 The 14 characters coded in columns 2 and 3 of table 2 (positions 21-27 and 31-37) represent combining marks which are non-spacing characters, that is, characters whose use is not followed by the forward movement of an output device. In a character string, these non-spacing characters are input before the characters they modify. Multiple combining marks associated with one letter are to be encoded in the order in which they appear, reading left to right or top to bottom. They are intended to be combined with other spacing characters in this International Standard or characters from the basic Cyrillic set. These combining marks (e.g. diaeresis) are used liberally in the non-Slavic languages that have Cyrillic-based alphabets. The BACKSPACE character (hexadecimal code 08 in ISO/IEC 646) should not be used when encoding non-spacing characters.

6.5 The rendering of graphic characters is intended solely to identify uniquely the additional Cyrillic script letters used by non-Slavic languages. The graphics used do not necessarily represent the most desirable calligraphic forms.

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

Non-Slavic languages using Cyrillic script characters from this International Standard

Tables A.1 and A.2 show the usage by non-Slavic languages of Cyrillic script characters defined in this International Standard. Only lower-case letters are shown. Variant graphics are enclosed in parentheses. Table A.1 is a listing by non-Slavic language, presented alphabetically according to their English-language names. Table A.2 is a listing of non-spacing characters and identifies with which non-Slavic language each is used. Characters are presented in ascending order according to the codes assigned in this International Standard.

Table A.1 — Listing by non-Slavic language[illegible]

Table A.1 — (concluded)

Language	Characters and combined forms	Combining marks
Khanty (Surgut)	ä ɟ ɥ ɵ ẽ ỹ ə	̈́ ̈́ ' (apostrophe)
Khanty (Vakh)	ä ɟ ɥ ɵ ẽ ỹ ə ẽ	̈́ ̈́ ' (apostrophe)
Kirghiz	ɥ ɵ ɣ	̈́
Komi	i ẽ	̈́
Komi (Molodtsov)	d ɬ ʒ ɹ ʒ j' ɭ (ɭ) ɮ (ɮ) ɵ ɕ ɮ	̈́ ̈́ ' (apostrophe)
Komi (Permyak)	i ɵ	̈́
Koryak	ɟ ɥ	̈́ ' (apostrophe)
Kurdish	ṛ (ṛ') ə ə' (ə') k̇ (k') ǫ̇ (p') ṛ (ṛ') ṛ (ṛ') ɮ (ɮ') ɮ (ɮ') ɮ (ɮ') q w	̈́ ̈́ ' (apostrophe)
Lak	I	
Lezghian	I	
Lithuanian		' (apostrophe)
Mansi	ɥ (ɮ')	̈́ ' (apostrophe)
Mari (Meadow)	ä ɮ ɵ ỹ	̈́
Mari (Mountain)	ä ɮ ɵ ỹ ɮ	̈́
Moldavian	ʒ	̈́ ' (apostrophe)
Mongolian	ɵ ɣ	
Mordvin (Erzya)	ɮ	
Mordvin (Moksha)	æ ẽ ɭ ɮ ɵ ɣ ɮ ẽ	̈́ ̈́
Nenets	ɥ	̈́ ' (apostrophe)
Nivkh		' (apostrophe)
Ossetic (Digor)	æ	' (apostrophe)
Ossetic (Iron)	æ	
Romany	ṛ ɮ	̈́
Sami	ɥ	̈́ ' (apostrophe)
Selkup	ä ɬ	̈́ ' (apostrophe)
Shor	i ɮ ɵ ỹ	̈́
Tabasaran	I	
Tajik	ɮ ɮ ɟ ɟ ɟ ɟ	̈́ ̈́
Tat	I	
Tatar	ʒ ɟ ɵ ɣ ɮ ə	̈́
Tatar (Kryashen)	ä ɮ ɵ ỹ	̈́
Turkmen	ʒ ɮ ɵ ɣ ə	̈́
Tuva	ɟ ɵ ɣ	̈́
Udekhe		' (apostrophe)
Udmurt	ʒ ẽ ɮ ɵ ɟ	̈́
Uighur	ɮ ʒ ɟ (ɟ) ɟ ɵ ɣ ɮ ə	̈́ ̈́
Uzbek	ɮ ɟ (ɟ) ỹ (ỹ) ɟ (ɟ)	̈́ ̈́ ̈́ ̈́
Yakut	ä ɮ i ɮ ɵ ẽ ỹ ɟ ɮ	̈́ ̈́

Table A.2 — Listing by non-spacing character

Character	Examples	Languages
COMBINING ACUTE	á é í ý á é ć ń ć	vowels: most languages consonants: Lezghian
COMBINING BREVE	ă ě ǔ ǖ ǘ ǚ ǜ	Chechen, Chuvash, Dungan, Eskimo, Kara-Kalpak, Moldavian, Mordvin, Romany, Uzbek
COMBINING CARON	ě ř	Chuvash
COMBINING CEDILLA	ç ç	Bashkir, Chuvash
COMBINING DIAERESIS	ä ë ÿ ö ü ǝ Ǟ ǟ Ǡ ǡ	Altai, Chuvash, Even, Gagauz, Kalmyk, Khakass, Khanty, Komi, Mari, Mordvin, Shor, Tatar, Udmurt
COMBINING DOUBLE ACUTE	ŷ	Chuvash
COMBINING GRAVE	è ù	Komi, Komi (Permyak), Khanty (Vakh)
COMBINING HIGH COMMA	ṛ ṡ ṣ ṭ ṣ ṭ ṣ ṭ	Kurdish
COMBINING LEFT DESCENDER	ҕ	Khakass
COMBINING LEFT OGONEK	ҕ Һ Ң Ҥ Ҧ ҧ	Abkhasian, Chechen, Chukchi, Evenki, Kazakh, Khanty, Koryak, Lapp, Mansi, Nenets, Tatar, Uighur
COMBINING MACRON	ā ī ū	Tajik, Yakut
COMBINING OGONEK	ҕ Җ Ң Ҥ Ҧ	Abkhasian, Chechen, Dargwa, Dungan, Kara-Kalpak, Uzbek, Yakut
COMBINING RIGHT DESCENDER	Җ Ң Ҥ Ҧ ҧ Ҩ	Abkhasian, Bashkir, Chechen, Chuvash, Dungan, Kalmyk, Kara-Kalpak, Kazakh, Khakass, Kirghiz, Tajik, Tatar, Turkmen, Uighur, Uzbek
COMBINING RING ABOVE	å	Selkup