Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Международная организация по стандартизации

Doc Type:	**Working Group Document**
Title:	**Additional Cyrillic characters for the UCS**
Source:	**Michael Everson**
Status:	**Working Document**
Date:	**1998-05-25**

This document proposes several Cyrillic characters for inclusion in the UCS, and discusses the rationale for their inclusion.

Most of these characters derive from ISO/TC46/SC4 standard ISO 10754:1996. The first three characters proposed are shown in the ALA/LC Romanization Tables. The last ten characters proposed in N1590 circulated on 1997-06-09 are shown the code table here, but the proposal summary form for them is in N1590.

The characters proposed to be added here are given below, with proposed code positions. The table is found on the next page, and the proposal summary form is appended thereafter.

hex	Name
0487	CYRILLIC TEN THOUSANDS SIGN
0488	CYRILLIC HUNDRED THOUSANDS SIGN
0489	CYRILLIC MILLIONS SIGN
04C5	CYRILLIC CAPITAL LETTER CHECHEN KA
04C6	CYRILLIC SMALL LETTER CHECHEN KA
04C9	CYRILLIC CAPITAL LETTER CHUVASH NG
04CA	CYRILLIC SMALL LETTER CHUVASH NG
04CD	CYRILLIC CAPITAL LETTER KOMI NG
04CE	CYRILLIC SMALL LETTER KOMI NG
04EC	CYRILLIC CAPITAL LETTER SELKUP OE
04ED	CYRILLIC SMALL LETTER SELKUP OE
04F6	CYRILLIC CAPITAL LETTER AISOR EL
04F7	CYRILLIC SMALL LETTER AISOR EL
04FA	CYRILLIC CAPITAL LETTER KURDISH QA
04FB	CYRILLIC SMALL LETTER KURDISH QA
04FC	CYRILLIC CAPITAL LETTER KURDISH WE
04FD	CYRILLIC SMALL LETTER KURDISH WE
04FE	CYRILLIC CAPITAL LETTER YA IE
04FF	CYRILLIC SMALL LETTER YA IE
0500	CYRILLIC CAPITAL LETTER KOMI DE
0501	CYRILLIC SMALL LETTER KOMI DE
0502	CYRILLIC CAPITAL LETTER KOMI DJE
0503	CYRILLIC SMALL LETTER KOMI DJE
0504	CYRILLIC CAPITAL LETTER KOMI DZE
0505	CYRILLIC SMALL LETTER KOMI DZE
0506	CYRILLIC CAPITAL LETTER KOMI ZJE
0507	CYRILLIC SMALL LETTER KOMI ZJE
0508	CYRILLIC CAPITAL LETTER YAKUT I WITH STROKE
0509	CYRILLIC SMALL LETTER YAKUT I WITH STROKE
050A	CYRILLIC CAPITAL LETTER JE WITH STROKE
050B	CYRILLIC SMALL LETTER JE WITH STROKE
050C	CYRILLIC CAPITAL LETTER KOMI ELJ
050D	CYRILLIC SMALL LETTER KOMI ELJ
050E	CYRILLIC CAPITAL LETTER EL WITH MIDDLE HOOK
050F	CYRILLIC SMALL LETTER EL WITH MIDDLE HOOK
0510	CYRILLIC CAPITAL LETTER MORDVIN EL KA
0511	CYRILLIC SMALL LETTER MORDVIN EL KA
0512	CYRILLIC CAPITAL LETTER EN WITH MIDDLE HOOK
0513	CYRILLIC SMALL LETTER EN WITH MIDDLE HOOK
0514	CYRILLIC CAPITAL LETTER ER KA
0515	CYRILLIC SMALL LETTER ER KA
0516	CYRILLIC CAPITAL LETTER KOMI ESJ
0517	CYRILLIC SMALL LETTER KOMI ESJ
0518	CYRILLIC CAPITAL LETTER KOMI TJE
0519	CYRILLIC SMALL LETTER KOMI TJE
051A	CYRILLIC CAPITAL LETTER EL WITH DESCENDER
051B	CYRILLIC SMALL LETTER EL WITH DESCENDER
051C	CYRILLIC CAPITAL LETTER ER WITH TICK
051D	CYRILLIC SMALL LETTER ER WITH TICK
051E	CYRILLIC CAPITAL LETTER SHORT I WITH DESCENDER
051F	CYRILLIC SMALL LETTER SHORT I WITH DESCENDER
0520	CYRILLIC CAPITAL LETTER EM WITH DESCENDER
0521	CYRILLIC SMALL LETTER EM WITH DESCENDER
0522	CYRILLIC CAPITAL LETTER E WITH DIAERESIS
0523	CYRILLIC SMALL LETTER E WITH DIAERESIS

TABLE xxx - Rows 04-05: CYRILLIC

	048	049	04A	04B	04C	04D	04E	04F	050	051	052
0	Ҫ	Ґ	Қ	Ұ	Ӏ	Ӑ	Ӡ	Ӱ	ԁ	Ӆ	Ԧ
1	ҫ	ґ	қ	ұ	Ӂ	ӑ	ӡ	ӱ	ԁ	ӆ	ԧ
2	Ӿ	Ғ	Ҥ	Ҳ	ӂ	Ӓ	Ӣ	Ӳ	ԃ	Ԣ	Ӭ
3	҃	ғ	ҥ	ҳ	Ӄ	ӓ	ӣ	ӳ	ԃ	ԣ	ӭ
4	҄	Ҕ	Ң	Ҵ	ӄ	Ӕ	Й	Ҹ	Ҙ	Ԥ	
5	҅	ҕ	ң	ҵ	Ҟ	ӕ	й	ҹ	ҙ	ԥ	
6	҆	Җ	Ҧ	Ҷ	ҟ	Ӗ	Ӧ	Ӏ	Ҙ	Ҁ	
7	҇	җ	ҧ	ҷ	Ҥ	ӗ	ӧ	ӏ	ҙ	ҁ	
8	҈	Ҙ	Ҩ	Ҹ	ҥ	Ә	Ѳ	Ы	Ӿ	Ҭ	
9	҉	ҙ	ҩ	ҹ	ә	ә	ѳ	ы	ӿ	ҭ	
A		Қ	Ҫ	Һ	Ӈ	Ӛ	Ӫ	Ԛ	Ԅ	Ԛ	
B		қ	ҫ	һ	ӈ	ӛ	ӫ	ԛ	ԅ	ԛ	
C		Қ	Ҭ	Ҍ	Ҷ	Ӝ	Ӕ	Ԝ	Ԉ	Ԗ	
D		қ	ҭ	ҍ	Ҏ	ӝ	œ	ԝ	ԉ	ԗ	
E		Ҝ	Ү	Ҍ	ҍ	Ӟ	Ӯ	Ҥ	Ԓ	Й	
F		ҝ	ү	ҍ		ӟ	ӯ	ҥ	ԓ	й	

A. Administrative

1. Title
Additional Cyrillic characters for the UCS
2. Requester's name
Michael Everson, EGT (WG2 member for Ireland)
3. Requester type
Expert contribution
4. Submission date
1998-05-25
5. Requester's reference
6a. Completion
This is a complete proposal.
6b. More information to be provided?
No

B. Technical -- General

1a. New script? Name?
No.
1b. Addition of characters to existing block? Name?
Yes. Cyrillic
2. Number of characters
52 (42 + 10)
3. Proposed category
Category A
4. Proposed level of implementation and rationale
Level 1 noncombining characters.
5a. Character names included in proposal?
Yes
5b. Character names in accordance with guidelines?
Yes
5c. Character shapes reviewable?
Yes
6a. Who will provide computerized font?
Michael Everson, Everson Gunn Teoranta
6b. Font currently available?
Yes
6c. Font format?
TrueType
7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?
Yes. See ISO 10754:1996, and the ALA/LC Romanization Tables.
7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?
Hardcopy is provided for WG2 distribution.
See http://www.indigo.ie/egt/standards/iso10646/pdf/iso-10754.pdf
8. Does the proposal address other aspects of character data processing?
No.

C. Technical -- Justification

1. Contact with the user community?

Yes: TC46/SC4, U.S. Library of Congress.

2. Information on the user community?

Librarians

3a. The context of use for the proposed characters?

Early 20th-century non-Slavic Cyrillic. Compatibility with ISO 10754:1996

3b. Reference

See ISO 10754:1996.

4a. Proposed characters in current use?

Yes

4b. Where?

Implementations of ISO 10754:1996

5a. Characters should be encoded entirely in BMP?

Yes.

5b. Rationale

Contemporary use and keeping them with the other Cyrillic characters.

6. Should characters be kept in a continuous range?

No.

7a. Can the characters be considered a presentation form of an existing character or character sequence?

No.

7b. Where?

7c. Reference

8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?

Some characters borrowed from the Latin alphabet into the Cyrillic alphabet have a similar appearance as their Latin originals. They function differently, however, in terms of pronunciation, sorting, etc.

8b. Where?

8c. Reference

9a. Combining characters or use of composite sequences included?

CYRILLIC LETTER E WITH DIAERESIS (see N1590) could be decomposed to E and COMBINING DIAERESIS.

9b. List of composite sequences and their corresponding glyph images provided?

See 9a and ISO/IEC JTC1/SC2/WG2 N1590.

10. Characters with any special properties such as control function, etc. included?

No

D. SC2/WG2 Administrative

To be completed by SC2/WG2

1. Relevant SC 2/WG 2 document numbers:

2. Status (list of meeting number and corresponding action or disposition)

3. Additional contact to user communities, liaison organizations etc.

4. Assigned category and assigned priority/time frame

Other Comments