1998-11-24

Universal Multiple-Octet Coded Character Set International Organization for Standardization Organisation Internationale de Normalisation Международная организация по стандартизации

Doc Type: Working Group Document

Title: Revised proposal for encoding the Glagolitic script in the UCS

Source: Michael Everson, EGT (IE)

Status: Expert Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 1998-11-24

This document is based on the proposal written by Joe Becker and published in UTR#3, and the proposal written by me in N1659. It is a revision of N1659, and contains the proposal summary.

A. Administrative

1. Title

Revised proposal for encoding the Glagolitic script in the UCS.

2. Requester's name

Michael Everson, EGT (WG2 member for Ireland).

3. Requester type

Expert contribution.

4. Submission date

1998-11-24.

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?

Yes. Glagolitic.

1b. Addition of characters to existing block? Name?

No.

2. Number of characters

92.

3. Proposed category

Category B.1.

4. Proposed level of implementation and rationale

As an alphabetic script, Glagolitic requires Level 1.

5a. Character names included in proposal?

Yes.

5b. Character names in accordance with guidelines?

Yes.

5c. Character shapes reviewable?

Yes (see below).

1

6a. Who will provide computerized font?

Michael Everson.

6b. Font currently available?

Yes.

6c. Font format?

TrueType.

7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?

Yes. ISO 6861:1996 is a coded character set for Glagolitic, as well as the bibliography below.

7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?

Yes, see annex A below.

8. Does the proposal address other aspects of character data processing?

No.

C. Technical – Justification

1. Contact with the user community?

Yes: Dr. Hinko Muren (Slovenia), ISO/TC46/SC2, ISO/TC46/SC4

2. Information on the user community?

Scholarly, ecclesiastical, and library communities.

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

Used to write Old Church Slavonic.

3b. Reference

See bibliography below.

4a. Proposed characters in current use?

Yes.

4b. Where?

By scholars, churchmen, and librarians.

5a. Characters should be encoded entirely in BMP?

Yes.

5b. Rationale

Accordance with the Roadmap, and close relationship of Glagolitic and Cyrillic texts.

6. Should characters be kept in a continuous range?

Yes, they should be encoded in a single block as presented here.

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

Yes: xx1E and xx4E could be represented by the sequences xx1D + xx09 and xx4D + xx39 respectively.

7b. Where?

7c. Reference

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

No.

8b. Where?

8c. Reference

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

Yes.

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

Generic combining characters can occur with Glagolitic base forms.

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

No.

E. Proposal

User community

Glagolitic was developed in the 9th century to write Old Church Slavonic (≈ Old Bulgarian). It arose more or less in parallel with the Cyrillic alphabet for the same language, and the two alphabets correspond to each other quite closely. The relationship between the origins of Glagolitic and Cyrillic is unknown, though St. Cyril is said to have had a hand in the creation of both. The Cyrillic script gradually supplanted Glagolitic, but Glagolitic continued in some liturgical use until the 19th century.

Structure

In the encoding, Glagolitic is treated as a separate script from Cyrillic, principally because the letter shapes are in most cases totally unrelated, with differences not at all arising from "mere font style". Glagolitic itself is seen in two slightly different styles, called the Bulgarian and the Croatian ("Illyrian" in Faulmann 1880). The Croatian form distinguishes uppercase and lowercase letters, although the difference in nearly all instances is merely one of size. The letterforms shown in the charts are Bulgarian style.

Directionality

Like Cyrillic, the Glagolitic script is written in linear sequence from left to right with no contextual modification of the letterforms.

Ordering

The ordering is basically the same as that of the (old) Cyrillic alphabet. Occasionally, sources show minor variations in the ordering of one or two characters (as they also do for Cyrillic). Ordering here follows that of Cubberly and Xaoypraeb.

Cubberly 1996 gives the following order and names:

azъ, buky, vědi/vědě, glagoli/glagolь, dobro, jestь/estъ, živěte, (d)zělo, zemlja, i/ižei, iže, ǵervь/dervь, kako, ljudъje/ljudije, myslite/myslěte, našь, onъ, pokoj, rьci, slovo, tvrdo/tverdo, ukъ/ikъ, frtъ, chěrъ/cherъ, otъ, ci, črvь, ša, štja, jerъ, jery, jerь, ětь/jatь, {jo}, [ju], jusъ malyi, jusъ malyi jotirovannyj, jusъ bolьšij, jusъ bolьšij jotirovannyj, (thita)/fita, ižica

Cubberly 1996 shows the Illyrian variant glyphs, but not for FRT and HER, and omits JO (which I have given in {curly brackets} here; I have given three characters for which he does not give names in [square brackets]).

Хабургаев 1986 gives the following order and names (spellings romanized here):

azŭ, buky, vědi, glagoli, dobro, jestŭ, živěte, ʒělo, zemlja, ižen, iže, g'ervĭ, kako, ljudije, myslite, našĭ, onŭ, pokoi, rĭci, slovo, tvrĭdo, oukŭ, frĭtĭ, xěrŭ, otŭ (omega), ci, črĭvĭ, ša, šta, jerŭ, jery, jerĭ, jatĭ, {jo}, [ju], jusŭ malyi, [je], jusŭ bolĭšii, [jo], fita, ižica

Хабургаев 1986 omits the Illyrian variants and omits JO (which I have given in {curly brackets} here; I have given three characters for which he does not give names in [square brackets]).

ISO 6861:1996 gives no explicit order and the following names:

azu buki, vedi, glagol´, dobro, esi´, živěte, zělo, zemlja, iže, i, g´erv´, kako, ljudi, myslite, naš, on, pokoj, r´ci, slovo, tverdo, uk, fert, xěr, omega, cy, červ´, ša, šta, er, ery, er´, jat´, jo, ju, as, es, jes, jas, fita, ižica

Faulmann 1880 gives the following order and names:

Az, Buki, Vědi, Glagol´, Dobro, Est´, Živěte, Zělo, Zemlja, Iže, I, Derv, Kako, Ljudi, Myslite, Naš, On, Pokoj, R´ci, Slovo, Tvr´do, Uk, Fert, Khěr, O, Šta, Tsi, Tšerv´, Ša, Jer, Jery, Jerek, Jet´, {Jo}, Ju, Ąs, Ęs, Jas, Jes, Thita, Ižica

Faulmann 1880 shows Illyrian variants but does not have the glyph for HER; he omits JO (which I have given in {curly brackets} here).

Leskien 1922 gives the following order (in romanization):

```
a, b, v, g, d, e, ž, dz, z, i, i, g, k, l, m, n, o, p, r, s, t, u, f, ch, o, št, c, č, š, ŭ, y, ĭ, ě, {jo}, ju, e, a, je, ja, f, ü
```

Leskien 1922 omits the Illyrian variants and omits JO (which I have given in {curly brackets} here).

Haarmann 1990 gives the following order (in romanization):

```
a, b, v, g, d, e, ž, dz, z, i, i, \gamma, k, l, m, n, o, p, r, s, t, u, f, \theta, x, \omega, št, c, č, š, ъ, y, ь, ě, {jo}, ju, ę, \varrho, je, j\varrho, \upsilon
```

Haarmann 1990 omits the Illyrian variants and omits JO (which I have given in {curly brackets} here).

Muren 1996 gives the following order and names:

```
azъ, buky, vědě, glagoli, dobro, jestъ, živěte, dzělo, zeml'a, iže, i, děrvь, kako, l'udьje, myslite, našь, onъ, pokojь, rьсі, slovo, tvrьdo, ukъ, frьtъ, hěrъ, otъ, šta, ci, črьvь, ša, jerъ, jery, jerь, jatь, {jo}, ju, ęsъ, jęsъ, osъ, josъ, ižica, fita
```

Muren 1996 shows the Illyrian variant glyphs, but not for FRT and HER, and omits JO (which I have given in {curly brackets} here.

In general it is the positions of CI, CHRV, SHA, and SHTA, as well as those of AS, ES, JAS, and JES, which differ from list to list.

Variant Glyph Forms

Several of the letters have variant glyph forms. These are not given separate character codes here. In ISO 6861:1996 five variant forms are explicitly encoded, and were in N1659. Discussion with the convener of ISO/SC4/WG2 indicated that in bibliographical contexts, probably *most* current use transliterates Glagolitic into Cyrillic in any case, and it is suggested that the variant characters in ISO 6861:1996 be unified with the basic letters at the present time. If in future evidence proves that the distinction must be made, there is room in the table to add them.

Only one character, GLAGOLITIC LETTER JO, is found in ISO 6861:1996 and not other sources to hand. Because ISO 6861:1996 presents it with its numeric value (7000) it is assumed that sources to hand in drawing up that standard did attest the character, and it is accordingly retained in this proposal.

Punctuation and diacritics

Glagolitic uses the following punctuation:

- , 002C COMMA
- . 002E FULL STOP
- ; 037E GREEK QUESTION MARK
- · 0387 GREEK ANO TELEIA
- : 0589 ARMENIAN FULL STOP
- : 10FB GEORGIAN PARAGRAPH SEPARATOR

ANO TELEIA is used, flanking a letter on either side, to show the numeric use of the letter (Faulmann 1880). Glagolitic uses numerous diacritical marks, many of them in common with Cyrillic. It is suggested

that research on the use of these in both Cyrillic and Glagolitic be initiated, and that any characters missing from the UCS be added at a future date to the general Combining Diacritical Marks block if necessary. It has been observed that a line above or a tilde above a letter or letters can be used to show the numeric use of the letter.

Numeric values

Glagolitic letters have inherent numeric values. XaбypraeB shows the characters with their oldest, Greek-based numeric values; the others show the later Slavonic values.

Name	ISO 6861	Cubberly	Хабургаев	Leskien	Faulmann	Diringer	Muren
AZ	1	1	1	1	1	1	1
BUKY	2	2		2	2	2	2
VEDI	3	3	2	3	3	3	3
GLAGOLI	4	4	3	4	4	4	4
DOBRO	5	5	4	5	5	5	5
JEST	6	6	5	6	6	6	6
ZHIVETE	7	7		7	7	7	7
DZELO	8	8	6	8	8	8	8
ZEMLJA	9	9	7	9	9	9	9
IZHE	20	20	8	20	10	10	10
I	10	10	10	10	20	20	20
GERV	_	30		30	30	30	30
KAKO	40	40	20	40	40	40	40
LJUDIJE	50	50	30	50	50	50	50
MYSLITE	60	60	40	60	60	60	60
NASH	70	70	50	70	70	70	70
ON	80	80	70	80	80	80	80
POKOJ	90	90	80	90	90	90	90
RCI	100	100	100	100	100	100	100
SLOVO	200	200	200	200	200	200	200
TVRDO	300	300	300	300	300	300	300
UK	400	400	(400)	400	400	400	400
FRT	500	500	500	500	500	500	500
HER	6000/600	600	600	600	600	600	600
OT	700	700	800	700	700	700	700
CI	900	900	900	900	900	900	900
CHRV	1000	1000	90	1000	1000	1000	1000
SHA	2000	(800)					(2000)
SHTA	800	(800)		800	800	800	800
JER	4000						(3000)
JERY							
JERI	3000						(?)
JAT	5000					800	800
JO	7000						
JU	8000						(5000)
AS			(900)	900		900	
ES	9000						
JAS	_						
JES	_						
FITA	9/500		9	9		9	
IZHICA	400		400	400			

Character sequence

The letter JERY could be considered as the sequence of JER and I. Indeed, the Russian JERU is also a sequence of SOFT SIGN and BYELORUSSIAN-UKRAINIAN I. As sources give GLAGOLITIC JERY in the repertoires, and as it relates to a unitary Cyrillic letter, it is suggested that it be encoded here.

Unicode Character Properties

Spacing letters, category "Lo", bidi category "L" (strong left to right) xx00-xx28, xx30-xx58

Bibliography

Cubberly, Paul. 1996. "The Slavic alphabets", in Peter T. Daniels and William Bright, eds. *The world's writing systems*. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0

Diringer, David. 1996. *The alphabet: a key to the history of mankind*. Munshiram Manoharlal. ISBN 81-215-0748-0

Faulmann, Carl. 1990 (1880). *Das Buch der Schrift*. Frankfurt am Main: Eichborn. ISBN 3-8218-1720-8 Haarmann, Harald. 1990. *Die Universalgeschichte der Schrift*. Frankfurt: Campus. ISBN 3-593-34346-0

ISO 6861:1996. Information and documentation — Glagolitic alphabet coded character set for bibliographic information interchange.

Leskien, A. 1922. *Handbuch der altbulgarischen (altkirchenslavischen) Sprache*. 6. Auflage. Heidelberg: Carl Winters Universitätsverlag.

Muren, Hinko. 1996-12-22. Ljubljana, Slovenia. Personal communication.

Nakanishi, Akira. 1990. Writing systems of the world: alphabets, syllabaries, pictograms. Rutland, VT: Charles E. Tuttle. ISBN 0-8048-1654-9

Unicode Consortium. 1992. Unicode Technical Report #3: exploratory proposals.

Хабургаев, Георгий Александрович. 1986. Старославянский язык. Москва: Просвещение.

Annex A

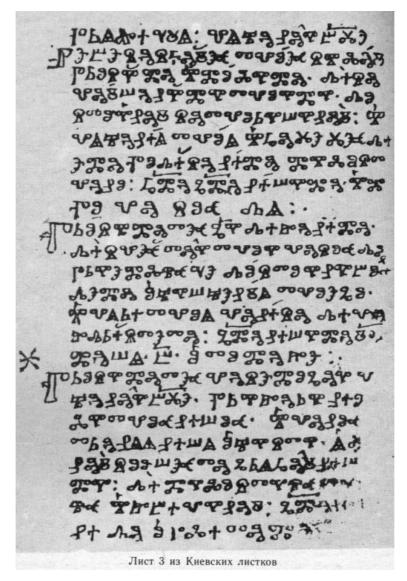
Sample from Leskien 1922:212.

Aus dem Codex Assemanianus.

Johannes IV, 5-42.

№ эгэ [уьаже] учгваэ баява уч жььан бажьноварь елеть влажь. 6. бу жа ма вивизточе вниче. Всяве жа WISSAGE RE SWG FEWT RAGALUS WILLS FI RWBGSFGVE. SSGTF1 жэ <u>ша арэ шэрм</u>ат. 7. 8 грдоэ жэлт эмф бүшүрдөг гэарычгамф эъэ эшват шаньж чи бынии, ан шыншигн ржг€пи. 9. банбэан эжэ жэгн гнжнручин төм сүнү түрүн жагы жэгэ ምምም ፠3₽6ም Ձት**ሜተ**ልል₽68₽€ <mark>2988€; ₽3 የ</mark>58ት4214€₩6 ២3 **2**€ 8₽**6**38 8434464364. 10. ОТАЧАТ 88384 8 6343 38. 473 28 ЧАЛАА4 ልተьራ ሮ3፠ዋ8 8 ነመን 32መሪ ኤ<mark>ልተ</mark>እያ<mark>ል€8 **መዋ**፣ ልተ</u>፠ልፊ **જ**ዋ **የዋመ**ዋ, መ**ሪ**ዋ</mark> шт гызатан ж гэхэ, 8 анав их шт чэлж хтчж. 11. ханхэжн 333 x344. 8321308, 42 134P414442p4 224m4, 2 30399343436 3304 ължень эта нела вен втит чиле крче; 12. элн тав Ľቌሑዋ8 **ኃደ8 ቌመፅ**೪ት **ፆትሠ**ኃኤኔ 8ትኦኔዮት, **ዋ**፠ኃ ሴትደመቆ **ፆት**፠ቆ ደመ<u></u>ውል**ን**ፆን****6 26, 8 we [8]& 13h9 17we 8 2etrovo 3h9 8 2bout 3h9; 13. 0we-жлэм[в] № г+1-вт, 14. + жжэ гтэмв эмв чэлвт, эсжэ +ыв л+жв omæ, fo 85% ተመብ የብጽብ€ቤተመዋ ደ€ የፅ የልኩብ, fa የout, 46% ተልብ -ንፉብ ልልደፁህ ይቀላይ እትጀትር መደር ይለ ውስር ይህ ውስር ይህ ውስር ውስር ውስር штезе че жүчэте чачегетт. 15. жанжын № гэтв жэгн жэежду жоженте те жожее те на жее узете на жее жее то тыпу жее ганьагния. 16. жанкаан эх хажая хап х гьхханат жэсжн ачээкэ

Note the use of COMMA, FULL STOP, GREEK QUESTION MARK, and GREEK ANO TELEIA here.



Note the use of the ARMENIAN FULL STOP, the GEORGIAN PARAGRAPH SEPARATOR and the GREEK ANO TELEIA here.

							Uncial		70	<u> </u>	- T	71	1
Uncial Greek	Cyrillic	Bulgarian Glago	Croat.	Names of the letters	Phonetic Val		Greek	Cyrillic	Bulgarian Glagoli	tsa	Names of the letters	Phonet.	ue.
A	11 4	ナま巾		az	a		T	Т	an an an	00 OO	turdo	t	300
ЬВ	5 -	* E	шш	buky	b	2	8 00	or v	8 3 3 3 E	EE 48	uk	и	400
li i	Ь	~ W		vedi	v	3	ф	φ. ‡	490	Ф	}ert	8	500
Γ	Г	222	90%	glagol	9	4	0			.	thrta	th	9
A	\triangle	Bar III	пь ОЪ	dobro	d	5	χ	X	ふるん	46 0 U	kKer ot	kh ō	6 <i>0</i> 0 7 0 0
6	ϵ			ost.	e-4	2 6	ω_{i}	w	මය ද	Q Y	06	U	700
	_	尹多子	Э Э	est',		۱ ۱	CU (50) Sem. sh	Ш	ШWW	Ш	sha	sh	-
T(TO)	* *	*	94 QQ	zhivete	zh	7	4 (662)	штЧ	<i>ጽ</i> ሕ ሉ	W W	shta	sht	800
3 (%:	5 \$	♣ ♣	占数	zelo	z-dz	8	(G,CZ,662) Sem <u>ts</u> ?	4	VVV	W W	tsi	ts	900
Z	3	Do Do	<i>8</i> 000	zemlya	Z	9	(C, TO 2) (T sem. ts!	Y	\$748	母	ch'ru	ch	1000
H	Н	T 35 4	双空	izhe (ije)	\bar{i}	10	8/02)	B	मा सिल्य		yer	ŭ-y	-
l ï	ïı	an SR	a X	i	ľ	20	21(0EtL)	3107	₩%		yery	y	-
	, ,	WK M	TPM	dew')	d'-d8	1 i	6(8)	Ь	ન્શ-8	TI	yerek	ĭ	-
Tr (24)	ኽ	مریم مر ده که	H	yew' b	t'-t3	30	64	古相	& BE B	BA	yat'- yet	ya-ye	800
K	K	334	5 5	kako	k	40	10	Ю	ملا مع	邓四	1	yu	-
人	λл	ኤዱ&	மிக	lyudy	2	50	18	Æ	-	_	ye	ye	-
M		æz11		myslite	m	60	个个(w)	ATÎ	怀	_	ens	eng	900
N		18 J L	P	nash	n	70	0 0+ LV	*************************************	* *	-	025	ong	-
0	0	8 8 8 8 B	-	on	0	80	I+ A	ЬА	ૠ ૠ	_	ye"s	yeng	-
П	П	क्षा कु		pokoy	p	90	I+ X	Ѭ	फु र %€	-	yo ⁿ s .	yong	-
Р	ρ	БЪБ Ь	ББ	r'tsi	r	100	238	五菱	_	-	ksi	ks	60
c	С	କ୍ଷ ଥ	1	slovo	5	200	* * * *	44	_	_	þsi	þs	700
				•)°	, 3	V	VY	₹ 8.	_	ızhitsa	y-v-i	

Fig. 212—The Early Slavonic alphabets (I)

Fig. 213-The Early Slavonic alphabets (II)

TABLE XX - Row xx: GLAGOLITIC

	xx0	xx1	xx2	xx3	xx4	xx5	xx6	xx7	
0	4	9	A	ተ	9	A			
1	Щ	₽º		Щ	Ф	☆			
2	v	Ь	P	જ	Ь	P			
3	%	8	3€	%	8	3€			
4	$\boldsymbol{\vartheta}$	W	£	δ	W	£			
5	Э	33	æ €	Э	æ	3%			
6	86	ф	€	*	ф	€			
7	ॐ	b	-0-	❖	b	⊕			
8	Øо	0	፟፟ቜ.	₽0	Q	₽.			G = 00 P = 00
9	8	V		82	V				
Α	8	쓩		8	성				
В	M	Ш		M	Ш				
С	þ	Ж		þ	W				
D	B	%		a	**				
E	னு	% %		ത്ര	48				
F	ተ	X		ተ	老				

TABLE XX - Row xx: GLAGOLITIC

dec	hex	Name	dec	hex	Name
000 001 002 003 004 005 006 007 008 009 010 011 012 013 014 015 016 017 018 020 021 022 023 024 025 026 027 028 030 031 032 033 034 035 036 037 038 039 040 041 042 043 044 045 046 047 048 049 050 061 061 070 071 072 073 074 075 076 077 078 079 070 071 072 073 074 075 076 077 078 079 070 071 072 073 074 075 076 077 078 079 070 071 072 073 074 075 076 077 078 079 070 071 072 073 074 075 076 077 078 079 077 078 079 070 071 072 073 074 075 076 077 078 079 079 070 071 072 073 074 075 076 077 078 079 079 070 071 072 073 074 075 076 077 078 079 079 079 070 071 072 073 074 075 076 077 078 079 079 079 079 079 079 079 079	001033456789000000000000000000000000000000000000	GLAGOLITIC CAPITAL LETTER AZ GLAGOLITIC CAPITAL LETTER BUKY GLAGOLITIC CAPITAL LETTER VEDI GLAGOLITIC CAPITAL LETTER VEDI GLAGOLITIC CAPITAL LETTER POBRO GLAGOLITIC CAPITAL LETTER DOBRO GLAGOLITIC CAPITAL LETTER DOBRO GLAGOLITIC CAPITAL LETTER DOBRO GLAGOLITIC CAPITAL LETTER ZHIVETE GLAGOLITIC CAPITAL LETTER REMAJA GLAGOLITIC CAPITAL LETTER KAKO GLAGOLITIC CAPITAL LETTER WISLITE GLAGOLITIC CAPITAL LETTER WASH GLAGOLITIC CAPITAL LETTER WASH GLAGOLITIC CAPITAL LETTER WASH GLAGOLITIC CAPITAL LETTER POKOJ GLAGOLITIC CAPITAL LETTER POKOJ GLAGOLITIC CAPITAL LETTER POKOJ GLAGOLITIC CAPITAL LETTER SUVO GLAGOLITIC CAPITAL LETTER POKOJ GLAGOLITIC CAPITAL LETTER FOR GLAGOLITIC CAPITAL LETTER SHA GLAGOLITIC CAPITAL LETTER WASH GLAGOLITIC SMALL LETTER WASH GLAGOLITIC SMALL LETTER WASH GLAGOLITIC SMALL LETTER WASH	089 090 091 092 093 094 095 096 097 100 101 102 103 104 105 1106 1107 1118 119 120 121 123 124 125 126 127	59 55 BC 55 EF 560 1 2 3 4 5 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 7	(This position shall not be used)