ISO/IEC JTC1/SC2/WG2 N2610

2003-08-24

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

L2/03-282

Doc Type: Working Group Document

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

Source: Michael Everson and Ralph Cleminson

Status: Expert Contribution

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

Date: 2003-08-24

This document is based on the proposal written by Joe Becker and published in UTR#3, and the proposal written by Michael Everson in N1659, revised in N1931. It is a revision of our N2555, and contains the proposal summary.

A. Administrative

1. Title

Final proposal for encoding the Glagolitic script in the UCS.

2. Requester's name

Michael Everson and Ralph Cleminson

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

4. Submission date

2003-08-24

5. Requester's reference (if applicable)

N1659, N1931, N2555.

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical -- General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

Yes.

Proposed name of script

Glagolitic.

1b. The proposal is for addition of character(s) to an existing block

Yes, three.

1b. Name of the existing block

General Punctuation.

2. Number of characters in proposal

97(94+3)

3. Proposed category (see section II, Character Categories)

Category B.1.

4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

As an alphabetic script, Glagolitic requires Level 1.

4b. Is a rationale provided for the choice?

No.

4c. If YES, reference

5a. Is a repertoire including character names provided?

Yes.

5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

5c. Are the character shapes attached in a legible form suitable for review?

Yes

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: Michael Everson, Fontographer.

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

Yes, see bibliography below.

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

Yes.

8. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes, see below.

9. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

Yes, see Unicode properties below.

C. Technical -- Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

Yes, exploratory drafts in N1659, N1931, N2555.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?

Ralph Cleminson, Dr Hinko Muren (Slovenia), ISO/TC46/SC2, ISO/TC46/SC4

2c. If YES, available relevant documents

N/A

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Scholarly, ecclesiastical, and library communities.

4a. The context of use for the proposed characters (type of use; common or rare)

Originally used to write Old Church Slavonic, and used in parts of medieval and early modern Croatia for both secular and ecclesiastical purposes.

4b. Reference

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

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

By scholars, churchmen, and librarians.

6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Ves

6c. If YES, reference

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

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

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

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No. Generic combining characters can occur with Glagolitic base forms.

12b. If YES, reference

13a. Does the proposal contain characters with any special properties such as control function or similar semantics?

13b. If YES, describe in detail (include attachment if necessary)

14a. Does the proposal contain any Ideographic compatibility character(s)?

No.

14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?

14c. If YES, reference

D. Proposal

User community

Glagolitic is considered to be the alphabet devised by St Cyril, probably in 862 CE, for his translation of the the Scriptures and liturgical books into Slavonic. The Glagolitic inscriptions and manuscripts which survive from the early period are relatively few in number, but of great philological importance. Amongst the Slavs of the Eastern Rite, Glagolitic was subsequently supplanted by the alphabet now known as Cyrillic (based on Greek uncials but following the order and function of the Glagolitic letters), which probably arose in late ninth-century Bulgaria. In those parts of Croatia where a vernacular liturgy was used, Glagolitic continued in use until modern times: the last Glagolitic missal was printed in Rome in 1893 with a second edition in 1905. In these areas Glagolitic is still occasionally used as a decorative alphabet.

Structure

In the encoding Glagolitic is treated as a separate alphabet from Cyrillic, both because of its historical primacy and because in almost every case the shapes of the letters in the two alphabets are completely dissimilar: the one can in no sense be regarded as a visual variant of the other. Glagolitic itself exists in two styles, known as round and square. Round Glagolitic is the original style and more geographically widespread (though surviving examples are less numerous); square Glagolitic (and the cursive style derived from it) was used in Croatia from the thirteenth century. There are a few documents written in a style intermediate between the two. The letter-forms used in the charts are round Glagolitic, since a number of characters found in this style are not represented in square Glagolitic, while the reverse (with

the exception of the last three characters in the table) is not the case.

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 of the Glagolitic alphabet is largely derived from that of the Greek alphabet, though nearly half the Glagolitic characters have no equivalent in Greek and not every Greek letter has its equivalent in Glagolitic. Scholars cannot recover the original (Cyrillo-Methodian) alphabet, there being no extant documents from this period; its content and order can nevertheless be reconstructed with reasonable certainty on the basis of later abecedaria and a number of acrostics of extremely early origin. It should however be recognised that the order of the alphabet as represented in the surviving documents is not entirely stable, particularly towards the end.

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.

Punctuation and diacritics

A number of punctuation characters are used in Glagolitic. Several of these are generic punctuation characters: COMMA, FULL STOP, SEMICOLON (= GREEK QUESTION MARK), and MIDDLE DOT (= GREEK ANO TELEIA) are used, and in edited texts quotation marks are used. Three unencoded punctuation marks have been observed, which have also been noted from time to time in Latin and other scripts. We believe they should be encoded in the General Punctuation block. They form a part of this proposal.

, 002C COMMA . 002E FULL STOP

; 037E (GREEK QUESTION MARK

· 0387 GREEK ANO TELEIA

: 0589 ARMENIAN FULL STOP

: 10FB GEORGIAN PARAGRAPH SEPARATOR

∴ 2056 TRIANGULAR COLON
 ∴ 2058 DIAMOND COLON
 ∴ 2059 QUINTUPLE COLON

The punctuation mark SEMICOLON 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. A number of additional punctuation marks are also required

Names

The names of the letters present a number of problems. The first is that regular phonological developments have caused some of them to diverge in the modem languages; **baye** črivi** CHRIVI, for example,

becomes črv in Croatian, upus črăv in Bulgarian and uepus červ in Russian. Secondly, the names of some letters used only in very early documents (such as & SPIDERY HA) have not survived. This particularly affects the nasal voweIs, all of which are known by the name jus (which corresponds to Slavonic peax josŭ Yusu), differentiated by various modern adjectives. For the letter even the phonetic value is uncertain: ö, jö or jo have been suggested. We have used Yo, which was also used to name this character in ISO 6861:1996. The names used here are in general use today, even though some of them are ancient. The letters have inherent numerical values. Those from 1 to 1000 are well attested; the values 2000, 3000 and 5000 for III SHA, & YERU and P YU respectively are supported by fifteenth-century evidence; the higher values assigned to some letters by modern authors are not reliable.

				Value	ISO 6861
ተ	₼₺₺₺	аzъ	AZU	1	1
巴	Ľ ቌ፞፞፞፞፞፞፞፞፞፞፞፞፞፟፟፟፟፟፟፟፟፟፟፟፟፟፟፟፟፟፟፟፟፟፟	buky	BUKY	2	2
v	የል &&	vědě	VEDE	3	3
%	% & +% 3& <u>%</u>	glagoli	GLAGOLI	4	4
$\boldsymbol{\delta}$	ѧӭшьэ	dobro	DOBRO	5	5
3	32WX	jestъ	JESTU	6	6
36	% ምያልመን	živěte	ZHIVETE	7	7
ॐ	&A&3	3ělo	DZELO	8	8
Θ_0	ል ቆ፠ ፥ ል	zemlja	ZEMLJA	9	9
ሞ	ሞ ፠ ን	iže	IZHE	10	20
82	% %3	iže	INITIAL IZHEI	10	20
8	8	i	I	20	10
M	маьчя	djervь	DJERVI	30	
Þ	ት ተ	kako	KAKO	40	40
æ	 &₽&ጟ϶	ljudьje	LJUDIJE	50	50
ஜ	ም & ድ Ձ& ድ መን	myslite	MYSLITE	60	60
ብ	የ ትШጸ	našь	NASHI	70	70
9	ን የ	опъ	ONU	80	80
₽°	ም 8ሩ ሮ ች	рокојь	POKOJI	90	90
Ь	ь፞፞፞፞፞፞ጜ፞፞፞፞፞፞ጜ፠	гьсі	RICI	100	100
8	ይጼይያ	slovo	SLOVO	200	200
00	առբաց	tvrьdo	TVRIDO	300	300
33	3 ₽4Æ	икъ	UKU	400	400
Φ	ФРЯШЯ	frьtъ	FRITU	500	500
b	БАЬ Х	hěrъ	HERU	600	6000/600
0	9W&	отъ	OTU	700	700
П			PE		
ሤ	ጸተ	šta	SHTA	800	800
V	V %	ci	CI	900	900
쓩	\$ ₽ጟ&ጟ	črьvь	CHRIVI	1000	1000
Ш	Шф	ša	SHA	2000	2000
×X	ЭЬЖ	јегъ	YERU	3000	4000
ব	ЭРД	јегь	YERI		3000
A	₩ 00 %	јаtь	YATI	800	5000
Ď			SPIDERY HA		
\mathcal{P}	P	ju	YU	5000	8000
€			SMALL YUS		_
Æ			SM. MASC. YUS		
A			YO		7000
3€			IOT. SMALL YUS		_
3€			BIG YUS		9000
3 €			IOT. BIG YUS		_
•	♦€ ℤℿͱ	fita	FITA		9/500
ጀ	ଅ ፠፠ የ	ižica	IZHICA		400

Unicode Character Properties

Spacing letters, category "Lu", bidi category "L" (strong left to right) 2C00-2C2E

Spacing letters, category "Ll", bidi category "L" (strong left to right) 2C30-2C5E

Bibliography

Cubberley, Paul. 1988. "On the Origin and Development of the Slavonic Letternames", in *Australian Slavonic and East European Studies*, 2, 29-54. ISSN 0818-8149.

Cubberley, 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

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.

Marti, Roland. 1999. "Abecedaria: a key to the original Slavic alphabet", pp. 175-200 of *Thessaloniki, Magna Moravia: Proceedings of the International Conference, Thessaloniki* 16-19 October 1997. Thessaloniki: Hellenic Association for Slavic Studies. ISBN 960-85959-3-2

Miklas, Heinz. 1998. "Griechisches Schriftdenken und slavische Schriftlichkeit. Glagolica und Kirillica zwischen Verschriftlichung", in Verschriftung und Verschriftlichung. Aspekte des Medienwechsels in verschiedenen Kulturen und Epopchen, edd. Ch. Ehler und U. Schaefer. ScriptOrialia 94. Tübingen. pp. 132-155.

Miklas, H., ed. 2000. *Glagolitica: zum Ursprung der slavischen Schriftkultur.* Unter der Mitarbeit von V. Sadovski und S. Richter. ÖAW. (Schriften der Balkankomission. Philos.-hist. Kl., Philolog. Abteilung; 41) Wien: Verlag der Österreichischen Akademie der Wissenschaften.

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.

Vajs, J. 1932. Rukověť hlaholské paleografie. Praha: Slovanský ústav.

Žubrinić, Darko. 1996. Hrvatska glagoljica. Zagreb: HDK sv. Jeronima. ISBN 953-6111-35-7

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

Ягич, Игнатий Викентьевич. 1911. *Глаголическое письмо*. Санктпетербург: Имп. Акад. Наук (Энциклопедия славянской филологии, 3, III).

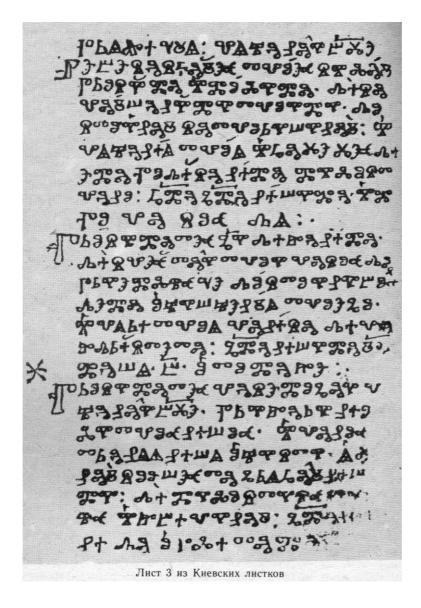
Annex A

Sample from Leskien 1922:212.

የተስዋህተንመብቼ ይመስተትል ጀደነትል የተብጀርት የተለያ መመ ይመውል የተስመር ጀተነት የተስዋህተንመብቼ ይመስተትል ጀደነትል የተለያ ተለያ ተለያ መመ ይመውልን ተላያ መተነት የተስዋህተንመብቼ ይመስተትል ጀደነትል የተለያ የመመስታ የተለያ መተነት የተስዋህተን መተነት የተስዋህተን መተነት የተስዋህተን መተነት የተስዋህተን መተነት የተስዋህተን መተነት የተስዋህተን መመን መመውስን ተለያ መመስታ የተለያ መመስታ መመን መመስታ የተለያ መመስታ የተለያ መመስታ የተለያ መመስታ የተለያ መመስታ የተለያ መመስታ የመመስታ የተለያ መመስታ የተለ

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

Sample from Хабургаев 1986:20.



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

TABLE XX - Row 2C: GLAGOLITIC

	2C0	2C1	2C2	2C3	2C4	2C5		205
0	ሑ	ቶ	X	4	P	-X	6	• •
1	Щ	9	A	Ш	9	A	8	•:•
2	v	₽	Ď	v	₽ º	\hat{Q}	9	::
3	%	Ь	P	%	Ь	P		
4	$\boldsymbol{\vartheta}$	8	Æ	B	R	æ		
5	Э	W	æ	Э	W	r⊲€		
6	86	3	*	*	3	*		
7	ॐ	ф	æ	♀	ф	3€		
8	Øо	b	æ	₿0	b	3€		G = 00 P = 00
9	Y	0	æ €	T	Q	æ€		
Α	8	П	-0-	82	П	-0-		
В	8	В	፟፟፟፟፟፟፟፟፟፟፟፟፟፟፟	8	W	₽.		
С	M	V	7	M	V	7		
D	þ	8	H	þ	쓩	H		
E	A	Ш	W	A	Ш	W		
F	कु	X		ரு	A X			

TABLE XX - Row 2C: GLAGOLITIC

hex	Name	hex	Name
00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F 11 12 13 14 15 16 17 18 19 11 11 11 11 11 11 11 11 11 11 11 11	GLAGOLITIC CAPITAL LETTER AZU GLAGOLITIC CAPITAL LETTER BUKY GLAGOLITIC CAPITAL LETTER BUKY GLAGOLITIC CAPITAL LETTER VEDE GLAGOLITIC CAPITAL LETTER GLAGOLI GLAGOLITIC CAPITAL LETTER DOBRO GLAGOLITIC CAPITAL LETTER ZHIVETE GLAGOLITIC CAPITAL LETTER ZHIVETE GLAGOLITIC CAPITAL LETTER ZHIVETE GLAGOLITIC CAPITAL LETTER ZEMLJA GLAGOLITIC CAPITAL LETTER IZHE GLAGOLITIC CAPITAL LETTER INITIAL IZHE GLAGOLITIC CAPITAL LETTER INITIAL IZHE GLAGOLITIC CAPITAL LETTER INITIAL IZHE GLAGOLITIC CAPITAL LETTER BERVI GLAGOLITIC CAPITAL LETTER BERVI GLAGOLITIC CAPITAL LETTER MYSLITE GLAGOLITIC CAPITAL LETTER MYSLITE GLAGOLITIC CAPITAL LETTER POKOJI GLAGOLITIC CAPITAL LETTER PORU GLAGOLITIC CAPITAL LETTER PORU GLAGOLITIC CAPITAL LETTER SHOU GLAGOLITIC CAPITAL LETTER PORU GLAGOLITIC CAPITAL LETTER FRITU GLAGOLITIC CAPITAL LETTER FRITU GLAGOLITIC CAPITAL LETTER PRE GLAGOLITIC CAPITAL LETTER PRE GLAGOLITIC CAPITAL LETTER SHA GLAGOLITIC CAPITAL LETTER SHOL GLAGOLITIC CAPITAL LETTER SHOL GLAGOLITIC CAPITAL LETTER SHI GLAGOLITIC CAPITAL LETTER SH	59 5A 5B 5C 5D 5E 5F 2056 2058 2059	RLAGOLITIC SMALL LETTER IOTATED BIG YUS GLAGOLITIC SMALL LETTER FITA GLAGOLITIC SMALL LETTER IZHITSA GLAGOLITIC SMALL LETTER SHTAPIC GLAGOLITIC SMALL LETTER TROKUSTASTI AZU GLAGOLITIC SMALL LETTER LATINATE MYSLITE (This position shall not be used) TRIANGULAR COLON DIAMOND COLON QUINTUPLE COLON
23 24 25 26 27 28 29 20 21 22 20 21 21 33 34 35 36 37 38 39 30 31 31 32 33 34 35 36 37 38 39 31 41 42 44 44 44 44 44 44 44 44 44 44 44 44	GLAGOLITIC CAPITAL LETTER JU GLAGOLITIC CAPITAL LETTER SMALL YUS GLAGOLITIC CAPITAL LETTER SMALL MASCULINE YUS GLAGOLITIC CAPITAL LETTER SMALL MASCULINE YUS GLAGOLITIC CAPITAL LETTER BIG YUS GLAGOLITIC CAPITAL LETTER FITA GLAGOLITIC CAPITAL LETTER FITA GLAGOLITIC CAPITAL LETTER SHTAPIC GLAGOLITIC CAPITAL LETTER SHTAPIC GLAGOLITIC CAPITAL LETTER TROKUSTASTI AZU GLAGOLITIC CAPITAL LETTER LATINATE MYSLITE (This position shall not be used) GLAGOLITIC SMALL LETTER AZU GLAGOLITIC SMALL LETTER BUKY GLAGOLITIC SMALL LETTER PEDE GLAGOLITIC SMALL LETTER BUKY GLAGOLITIC SMALL LETTER BUKY GLAGOLITIC SMALL LETTER CAGOLI GLAGOLITIC SMALL LETTER PESTU GLAGOLITIC SMALL LETTER ZHIVETE GLAGOLITIC SMALL LETTER ZHIVETE GLAGOLITIC SMALL LETTER ZHIVETE GLAGOLITIC SMALL LETTER INITIAL IZHE GLAGOLITIC SMALL LETTER NASHI GLAGOLITIC SMALL LETTER NOU GLAGOLITIC SMALL LETTER NOU GLAGOLITIC SMALL LETTER NOU GLAGOLITIC SMALL LETTER POKOJI GLAGOLITI		
4E 4F 50 51 52 53 54 55 56 57 58	GLAGOLITIC SMALL LETTER SHA GLAGOLITIC SMALL LETTER YERU GLAGOLITIC SMALL LETTER YERI GLAGOLITIC SMALL LETTER YATI GLAGOLITIC SMALL LETTER SPIDERY HA GLAGOLITIC SMALL LETTER SPIDERY HA GLAGOLITIC SMALL LETTER SMALL YUS GLAGOLITIC SMALL LETTER SMALL MASCULINE YUS GLAGOLITIC SMALL LETTER YO GLAGOLITIC SMALL LETTER BIG YUS GLAGOLITIC SMALL LETTER BIG YUS	Plane 00	Pow 2