ISO/IEC JTC1/SC2/WG2 N2555

2002-12-04

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

L2/02-448

Doc Type:Working Group DocumentTitle:Revised proposal for encoding the Glagolitic script in the UCSSource:Michael Everson and Ralph CleminsonStatus:Expert ContributionAction:For consideration by JTC1/SC2/WG2 and UTCDate:2002-12-04

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. It is a revision of Michael Everson's N1931, and contains the proposal summary.

A. Administrative

Title
 Revised proposal for encoding the Glagolitic script in the UCS.
 Requester's name
 Michael Everson and Ralph Cleminson
 Requester type (Member body/Liaison/Individual contribution)
 Individual contribution.
 Submission date
 2002-12-04
 Requester's reference (if applicable)
 Choose one of the following:
 6a. This is a complete proposal
 Yes.
 6b. More information will be provided later
 No
 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 No. 1b. Name of the existing block 2. Number of characters in proposal 88 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? Yes.

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 and N1931.

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?

Yes.

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? No.

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, 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 tenth-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. 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 stapic) 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 derived from that of Greek, though nearly half the Glagolitic characters have no equivalent in Greek and not every Greek letter has its equivalent in Glagolitic. The original (Cyrillo-methodian) alphabet is not recoverable, 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. The character ϵ SMALL YUS should be distinguished from ϵ (or ϵ) by means of VARIANT SELECTOR-1. It occurs sporadically in a few early manuscripts, but is clearly not a random variant, since it occurs only in masculine participles (in this case, though, it is worth considering that if it isn't a random variant it might be worth encoding). 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

Glagolitic uses the following punctuation:

- , 002С СОММА
- . 002E FULL STOP
- ; 037E GREEK QUESTION MARK
- · 0387 GREEK ANO TELEIA
- : 0589 ARMENIAN FULL STOP
- : 10FB Georgian paragraph separator

The punctuation mark 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.

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; 464898 *črivi* CHRIVI, for example, becomes *črv* in Croatian, *upъb črăv* in Bulgarian and *uepbb 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 Old Church Slavonic 4628 *josŭ* YUSU), differentiated by various modern adjectives. For the letter **a** 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 letters have inherent numerical values. Those from 1 to 1000 are well attested; the values 2000, 3000 and 5000 for u 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	ተ የ•ጽ	аzъ	AZU	1	1
Ľ	Ľ ንኦፚ	buky	BUKY	2	2
v	የ <mark>ል</mark> ቆል	vědě	VEDE	3	3
%	֍֎ֈֈ֍֎֎	glagoli	GLAGOLI	4	4
	ശാ ണ്ടാ	dobro	DOBRO	5	5
Э	ንደመዳ	jestъ	JESTU	6	6
ж	፠፠ሦልመን	živěte	ZHIVETE	7	7
₽	֎֎֎֎	3ĕlo	DZELO	8	8
Θ_0	֍℈ℛՅ֎	zemlja	ZEMLJA	9	9
ሞ	ሞಹን	iže	IZHE	10	20
8	<mark></mark>	iže	INITIAL IZHEI	10	20
8	8	i	Ι	20	10
M	маьюя	djervь	DJERVI	30	—
þ	ኦሑኦን	kako	KAKO	40	40
ቆ	ሌ ሥሌ ୟ ን	ljudьje	LJUDIJE	50	50
ញ	፠ፈጄይዔጄመን	myslite	MYSLITE	60	60
ተ	ዋቀጠል	našь	NASHI	70	70
9	ንቶል	onъ	ONU	80	80
P	ምፀኑ ንም	рокојь	POKOJI	90	90
Ь	<mark>ь</mark> ጜ v 宠	rьсі	RICI	100	100
8	ຬ Ⴥ ാ ഹാ	slovo	SLOVO	200	200
συ	መծրষษց	tvrьdo	TVRIDO	300	300
H	መ ኑ&	ukъ	UKU	400	400
φ	ቀ ь አመል	frьtъ	FRITU	500	500
b	рара	hěrъ	HERU	600	6000/600
0	900-X	otъ	OTU	700	700
ď			PE		
Ж	ጜቶ	šta	SHTA	800	800
V	v %	ci	CI	900	900
쓩	%Ь &&&&	črьvь	CHRIVI	1000	1000
Ш	Шф	ša	SHA	2000	2000
R	ЭЬ-Х	jerъ	YERU	3000	4000
Ł	эья	jerь	YERI		3000
A	AWA	jatь	YATI		5000
Q			SPIDERY HA		
ሆ	P	ju	YU	5000	8000
£			SMALL YUS		
☆			YO		7000
Æ			IOT. SMALL YUS	5	—
Æ			BIG YUS		9000
₩ €			IOT. BIG YUS		—
·Ð·	ው 密መቶ	fita	FITA		9/500
<u>8</u> .	<mark>፞</mark> ፝፞፞፝ፚቚ፝፞፞፞፝ጞኯኯ	ižica	IZHICA		400
]	<u></u> ይመቶሞ <mark>8</mark> ዶን	stapić	STAPIC		

Unicode Character Properties

Spacing letters, category "Lo", bidi category "L" (strong left to right) xx00-xx2B, xx30-xx5B

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
- 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.

Aus dem Codex Assemanianus.

Johannes IV, 5-42.

₩-8 эгэ [¥b&#€] ¥-8784.3 829824 ¥-8 %b+4.64 &+#+b3826 *и*трания страния с RATTO RUISTON 6. LA X3 NB 2080313V6 8449VA6. 82824 X3 жэ ша арэ шэвюнн. 7. 8 граз жэрн эте внаньтэс гэчьегное VILAT. SAISIAI IS SEBLE A ATAGE SS FTOT. 8. 34378VS C3 эка эшемт шальж ча кылаа, ал шылшага эжгепа. 9. калкала FTWT #376T 2+#+6478876 22586; 73 #68+4846608 23 26 80438 2437+baf3be. 10. OWEVAN+ 82328 8 b343 38. +83 28 VARAA+ альа шэжте е коэ зема калкэлсе от альала от гтот, мат **СР гызатан в гэхэ, 8 аная их ср узаж жруж. 11. канкэйн** 333 x324. 232L302 42 L34P-4L4454 224 224 224 2354 3504 3506 ължерья этя нела эез занше чэлж жече; 12. зан таз 29A78 328 906V+ F+U3%3 8+>3V+, TX3 A+204 F+884 208A3F3V8 26. 8 we [8]& P329 FTW-6 8 2-6TP3V3 323 8 21-9WT 323; 13. OW-6upartae 253524 p343 32. Asvie Lase 304 Asvel 5336 Ageve-9339, ₽9 538+08 9888€8+07 2€ 96 921-8, ₽8 998+, 4689 +68

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

Sample from Хабургаев 1986:20.

1º6Adot YVA: PATA _P32389 84483 P5389 23 88 9P384 m gr gr gr 800 gape 883004 g SP ACV 3-27 2+8 PAP2: 627327 35 89d va -0698***253**~742*****A +203-3+203. 148836009 PSP 7998 NE 93 2398998 8. 949214728 BA 988 2mg 28380, 30 · 9723 900 SOL M + 29234 V 5998-368-9 +9 f 94113d. poor 0069-PAA P+WA 30099. 2 13889994 9 7. Y: A+7 829900 8 (There 8+ AA 2100+0099 Лист 3 из Киевских листков

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

	2C0	2C1	2C2	2C3	2C4	2C5
0	ሐ	ዋ	Æ	ሐ	ዋ	R
1	Щ	9	А	ଅ	9	A
2	Ŷ	ሞ	Q	જ	ሞ	Û
3	%	Ь	ሆ	%	Ь	ሆ
4	ծ	8	£	ծ	8	£
5	Э	UU		Э	ω	会
6	Ж	39	Э€	ж	3	€
7	৵	φ	3£	Ŷ	φ	3 €
8	Øo	b	%€	Øo	b	∻€
9	ዋ	0	-0-	ሞ	Q	· 0 ·
A	8	Ъ	§ .	8	Ъ	&
В	8	Ч	7	8	ሤ]
С	M	V		M	V	
D	2	성		þ	舟	
E	Ֆ	Ш			Ш	
F	ஜ	R		ஜ	R	

TABLE XX - Row 2C: GLAGOLITIC

G = 00 P = 00

TABLE XX - Row 2C: GLAGOLITIC

dec	hex	Name	dec	hex	Name
000 001 002 003 004 005 006 007 008 010 011 012 013 014 015 016 017 018 020 021 022 023 024 025 026 027 028 029 030 031 032 033 034 035 036 037 038 039 040 041 042 033 034 035 036 037 038 039 040 041 022 023 024 025 026 027 028 029 030 031 032 033 034 035 036 037 038 039 040 041 042 043 044 045 056 057 058 059 066 067 058 059 066 067 078 055 056 057 058 059 066 067 077 078 077 078 077 078 080 081 082 083 084 085 066 077 078 077 078 080 081 082 083 084 085 086 087 088 087 088 088 088 088 088	001234566789ABCDEF01123456789ABCDEF01123456789ABCDEF0123456789ABCDEF01233456789ABCDEF012345456789ABCDEF01123456789ABCDEF01234556789ABCDEF01234556789ABCDEF0123456789ABCDEF01234556789ABC000000000000000000000000000000000000	GLAGOLITIC CAPITAL LETTER AZU GLAGOLITIC CAPITAL LETTER BUKY GLAGOLITIC CAPITAL LETTER VEDE GLAGOLITIC CAPITAL LETTER VEDE GLAGOLITIC CAPITAL LETTER DOBRO GLAGOLITIC CAPITAL LETTER DDBRO GLAGOLITIC CAPITAL LETTER ZHIVETE GLAGOLITIC CAPITAL LETTER ZHIVETE GLAGOLITIC CAPITAL LETTER ZHIVETE GLAGOLITIC CAPITAL LETTER INITIAL IZHE GLAGOLITIC CAPITAL LETTER INITIAL IZHE GLAGOLITIC CAPITAL LETTER INITIAL IZHE GLAGOLITIC CAPITAL LETTER INITIAL IZHE GLAGOLITIC CAPITAL LETTER NASHI GLAGOLITIC CAPITAL LETTER NINI GLAGOLITIC CAPITAL LETTER FIRI GLAGOLITIC CAPITAL LETTER FIRI GLAGOLITIC CAPITAL LETTER NINI GLAGOLITIC SMALL LETT	089 090 091 092 093 094 095 096 097 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126	59ABCDEF0123456789ABCDEF01234567789ABCDEF777777777777777777777777777777777777	GLAGOLITIC SMALL LETTER FITA GLAGOLITIC SMALL LETTER STAPIC (This position shall not be used) (This position shall not position shall not be used) (This position shall not be u