

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

L2/03-282R

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: 2003-09-04**

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, revised in N1931. It is a revision of our N2555, and contains the proposal summary. It corrects some errors in N2610.

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

2003-09-04

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?

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, 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?

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 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
;	003B	SEMICOLON
·	00B7	MIDDLE DOT
:	0589	ARMENIAN FULL STOP
∴	10FB	GEORGIAN PARAGRAPH SEPARATOR
∴	2056	TRIANGULAR COLON
∴	2058	DIAMOND COLON
∴	2059	QUINTUPLE COLON

The punctuation mark MIDDLE DOT 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 modern languages; **ѣриѣ** *črivi* CHRIVI, for example,

becomes *črv* in Croatian, *чрѣв чрѣвъ* in Bulgarian and *червь червь* in Russian. Secondly, the names of some letters used only in very early documents (such as **Ѡ** SPIDERY HA) have not survived. This particularly affects the nasal vowels, all of which are known by the name *jus* (which corresponds to Slavonic **Ѡ** *josŭ* YUSU), differentiated by various modern adjectives. For the letter **Ѡ** 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 not ancient. The letters have inherent numerical values. Those from 1 to 1000 are well attested; the values 2000, 3000 and 5000 for **Ш** SHA, **Ѧ** YERU and **Ѣ** YU respectively are supported by fifteenth-century evidence; the higher values assigned to some letters by modern authors are not reliable. Numeric values given in ISO 6861 which are incorrect are marked with an asterisk.

				Value	ISO 6861
Ѡ	ѠѠѦ	azъ	AZU	1	1
ѡ	ѡѡѦѦ	buky	BUKY	2	2
Ѣ	ѢѠѠѠ	vědě	VEDE	3	3
ѣ	ѣѠ+ѣѡѦѦ	glagoli	GLAGOLI	4	4
Ѥ	ѤѡѠѠѡ	dobro	DOBRO	5	5
ѥ	ѥѠѠѦ	jestъ	JESTU	6	6
Ѧ	ѦѦѢѠѠѠѡ	živěte	ZHIVETE	7	7
ѧ	ѧѠѠѡ	zělo	DZELO	8	8
Ѩ	ѨѡѦѦѠѠ	zemlja	ZEMLJA	9	9
ѩ	ѩѦѡ	iže	IZHE	10	20*
Ѫ	ѪѦѡ	iže	INITIAL IZHE	10	20*
ѫ	ѫ	i	I	20	10*
Ѭ	ѬѠѠѢѦѦ	djervъ	DJERVI	30	—*
ѭ	ѭѠѠѡ	kako	KAKO	40	40
Ѯ	ѮѢѠѠѠѡ	ljudъje	LJUDIJE	50	50
ѯ	ѯѦѦѦѠѠѠѠѠѡ	myslite	MYSLITE	60	60
Ѱ	ѰѠѠѦѦ	našъ	NASHI	70	70
ѱ	ѱѢѦѦ	onъ	ONU	80	80
Ѳ	ѲѡѡѡѦѦ	pokoјъ	POKOJI	90	90
ѳ	ѳѦѦѢѦѦ	řci	RICI	100	100
Ѵ	ѴѠѠѡѢѦѦ	slovo	SLOVO	200	200
ѵ	ѵѢѢѦѦѠѠѡ	tvrdо	TVRIDO	300	300
Ѷ	ѶѢѦѦѦ	ukъ	UKU	400	400
ѷ	ѷѢѦѦѦѦѦ	frъtъ	FRITU	500	500
Ѹ	ѸѠѠѦѦѦ	hěřъ	HERU	600	6000*/600
ѹ	ѹѠѦѦѦ	otъ	OTU	700	700
Ѻ			PE		
ѻ	ѻѠ	šta	SHTA	800	800
Ѽ	ѼѦѦ	ci	CI	900	900
ѽ	ѽѢѦѦѢѦѦ	čřъvъ	CHRIVI	1000	1000
Ѿ	ѾѠ	ša	SHA	2000	2000
ѿ	ѿѢѦѦ	jerъ	YERU	3000	4000*
Ѡ	ѠѢѦѦ	jerъ	YERI		3000*
ѡ	ѡѠѦѦѦ	jatъ	YATI	800	5000*
Ѣ			SPIDERY HA		
ѣ	ѣ	ju	YU	5000	8000*
Ѥ			SMALL YUS		—
ѥ			SM. YUS W. TAIL		
Ѧ			YO		7000*
ѧ			IOT. SMALL YUS		—
Ѩ			BIG YUS		9000*
ѩ			IOT. BIG YUS		—
Ѭ	ѬѦѦѦѠѠ	fita	FITA		9/500*

Ⓔ	ⒺⒻⒼⒿ	ižica	IZHICA	400*
Ј		štapić	SHTAPIC	
Ⓕ			TROKUSTASTI A	
Ⓜ			LAT. MYSLITE	

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

Punctuation, category “Po”, bidi category “CS”

2056, 2058-2059

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 albulgarischen (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 Verschriftung und Verschriftlichung”, in *Verschriftung und Verschriftlichung: Aspekte des Medienwechsels in verschiedenen Kulturen und Epochen*, edd. Ch. Ehler und U. Schaefer. (ScriptOralia; 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 Balkankommission. 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).

TABLE XX - Row 2C: GLAGOLITIC

hex	Name	hex	Name
00	GLAGOLITIC CAPITAL LETTER AZU	59	GLAGOLITIC SMALL LETTER IOTATED BIG YUS
01	GLAGOLITIC CAPITAL LETTER BUKY	5A	GLAGOLITIC SMALL LETTER FITA
02	GLAGOLITIC CAPITAL LETTER VEDE	5B	GLAGOLITIC SMALL LETTER IZHITSA
03	GLAGOLITIC CAPITAL LETTER GLAGOLI	5C	GLAGOLITIC SMALL LETTER SHTAPIC
04	GLAGOLITIC CAPITAL LETTER DOBRO	5D	GLAGOLITIC SMALL LETTER TROKUSTASTI AZU
05	GLAGOLITIC CAPITAL LETTER YESTU	5E	GLAGOLITIC SMALL LETTER LATINATE MYSLITE
06	GLAGOLITIC CAPITAL LETTER ZHIVETE	5F	(This position shall not be used)
07	GLAGOLITIC CAPITAL LETTER DZELO		
08	GLAGOLITIC CAPITAL LETTER ZEMLJA		
09	GLAGOLITIC CAPITAL LETTER IZHE		
0A	GLAGOLITIC CAPITAL LETTER INITIAL IZHE	2056	TRIANGULAR COLON
0B	GLAGOLITIC CAPITAL LETTER I	2058	DIAMOND COLON
0C	GLAGOLITIC CAPITAL LETTER DERVI	2059	QUINTUPLE COLON
0D	GLAGOLITIC CAPITAL LETTER KAKO		
0E	GLAGOLITIC CAPITAL LETTER LJUDIJE		
0F	GLAGOLITIC CAPITAL LETTER MYSLITE		
10	GLAGOLITIC CAPITAL LETTER NASHI		
11	GLAGOLITIC CAPITAL LETTER ONU		
12	GLAGOLITIC CAPITAL LETTER POKOJI		
13	GLAGOLITIC CAPITAL LETTER RITSI		
14	GLAGOLITIC CAPITAL LETTER SLOVO		
15	GLAGOLITIC CAPITAL LETTER TVRIDO		
16	GLAGOLITIC CAPITAL LETTER UKU		
17	GLAGOLITIC CAPITAL LETTER FRITU		
18	GLAGOLITIC CAPITAL LETTER HERU		
19	GLAGOLITIC CAPITAL LETTER OTU		
1A	GLAGOLITIC CAPITAL LETTER PE		
1B	GLAGOLITIC CAPITAL LETTER SHTA		
1C	GLAGOLITIC CAPITAL LETTER TSI		
1D	GLAGOLITIC CAPITAL LETTER CHRIVI		
1E	GLAGOLITIC CAPITAL LETTER SHA		
1F	GLAGOLITIC CAPITAL LETTER YERU		
20	GLAGOLITIC CAPITAL LETTER YERI		
21	GLAGOLITIC CAPITAL LETTER YATI		
22	GLAGOLITIC CAPITAL LETTER SPIDERY HA		
23	GLAGOLITIC CAPITAL LETTER JU		
24	GLAGOLITIC CAPITAL LETTER SMALL YUS		
25	GLAGOLITIC CAPITAL LETTER SMALL YUS WITH TAIL		
26	GLAGOLITIC CAPITAL LETTER YO		
27	GLAGOLITIC CAPITAL LETTER IOTATED SMALL YUS		
28	GLAGOLITIC CAPITAL LETTER BIG YUS		
29	GLAGOLITIC CAPITAL LETTER IOTATED BIG YUS		
2A	GLAGOLITIC CAPITAL LETTER FITA		
2B	GLAGOLITIC CAPITAL LETTER IZHITSA		
2C	GLAGOLITIC CAPITAL LETTER SHTAPIC		
2D	GLAGOLITIC CAPITAL LETTER TROKUSTASTI AZU		
2E	GLAGOLITIC CAPITAL LETTER LATINATE MYSLITE		
2F	(This position shall not be used)		
30	GLAGOLITIC SMALL LETTER AZU		
31	GLAGOLITIC SMALL LETTER BUKY		
32	GLAGOLITIC SMALL LETTER VEDE		
33	GLAGOLITIC SMALL LETTER GLAGOLI		
34	GLAGOLITIC SMALL LETTER DOBRO		
35	GLAGOLITIC SMALL LETTER YESTU		
36	GLAGOLITIC SMALL LETTER ZHIVETE		
37	GLAGOLITIC SMALL LETTER DZELO		
38	GLAGOLITIC SMALL LETTER ZEMLJA		
39	GLAGOLITIC SMALL LETTER IZHE		
3A	GLAGOLITIC SMALL LETTER INITIAL IZHE		
3B	GLAGOLITIC SMALL LETTER I		
3C	GLAGOLITIC SMALL LETTER DERVI		
3D	GLAGOLITIC SMALL LETTER KAKO		
3E	GLAGOLITIC SMALL LETTER LJUDIJE		
3F	GLAGOLITIC SMALL LETTER MYSLITE		
40	GLAGOLITIC SMALL LETTER NASHI		
41	GLAGOLITIC SMALL LETTER ONU		
42	GLAGOLITIC SMALL LETTER POKOJI		
43	GLAGOLITIC SMALL LETTER RITSI		
44	GLAGOLITIC SMALL LETTER SLOVO		
45	GLAGOLITIC SMALL LETTER TVRIDO		
46	GLAGOLITIC SMALL LETTER UKU		
47	GLAGOLITIC SMALL LETTER FRITU		
48	GLAGOLITIC SMALL LETTER HERU		
49	GLAGOLITIC SMALL LETTER OTU		
4A	GLAGOLITIC SMALL LETTER PE		
4B	GLAGOLITIC SMALL LETTER SHTA		
4C	GLAGOLITIC SMALL LETTER TSI		
4D	GLAGOLITIC SMALL LETTER CHRIVI		
4E	GLAGOLITIC SMALL LETTER SHA		
4F	GLAGOLITIC SMALL LETTER YERU		
50	GLAGOLITIC SMALL LETTER YERI		
51	GLAGOLITIC SMALL LETTER YATI		
52	GLAGOLITIC SMALL LETTER SPIDERY HA		
53	GLAGOLITIC SMALL LETTER YU		
54	GLAGOLITIC SMALL LETTER SMALL YUS		
55	GLAGOLITIC SMALL LETTER SMALL YUS WITH TAIL		
56	GLAGOLITIC SMALL LETTER YO		
57	GLAGOLITIC SMALL LETTER IOTATED SMALL YUS		
58	GLAGOLITIC SMALL LETTER BIG YUS		