

Proposal to Encode Some Outstanding Early Cyrillic Characters in Unicode

n3974

Yuri Shardt, Nikita Simmons, Aleksandr Andreev

In old, Slavic documents that come from Eastern Europe in the centuries between A.D. 1400 and 1700, it is possible to come across various unusual characters, whose forms have not yet entered the Unicode repertoire. As well, these symbols can occasionally be found in more recent publications by those Orthodox Christians (Old Believers) that do not accept the typographic changes introduced by Patriarch Nikon in the mid-1600s. The exact forms of these symbols often vary between different places and times. In the interest of creating a standard for faithfully typesetting Slavonic manuscripts, there is a need to include these symbols in Unicode. Since these symbols are often found in Slavonic Church books, these symbols will be encoded in the “Extended Cyrillic Block B” of the Unicode standard.

Characters

Table 1 presents a summary of the proposed characters for encoding in Unicode: Double O and Crossed O, which are found in many early Slavonic manuscripts. The double *o* is used in the words **двое** (two), **обо** (both), **обанадесять** (twelve), and **двоюнадесять** (twelve), where the bold letters denote the usual placement of the proposed double *o*. This letter would complement the other forms of *o* that are already present in the Unicode standard, including the monocular *o*, the binocular *o*, the double monocular *o*, and the multiocular *o*. However, unlike the ocular *o*'s, which are primarily used in different words denoting *eye*, the double *o* is used in words that denote *two*. As well, it can be noted that it never has any dots inside the *o*'s. The crossed *o* is primarily used in the word **окрест** (around, in the region of) in early Slavonic manuscripts. This character's usage is more common than that of the multiocular *o* and hence should also be added. Examples of both characters can be found in Figure 1 and Figure 2.

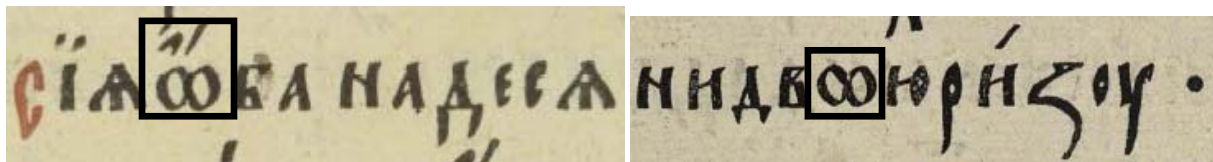


Figure 1: Extract from a 1553/4 Gospel showing the use of the Cyrillic double o (boxed).

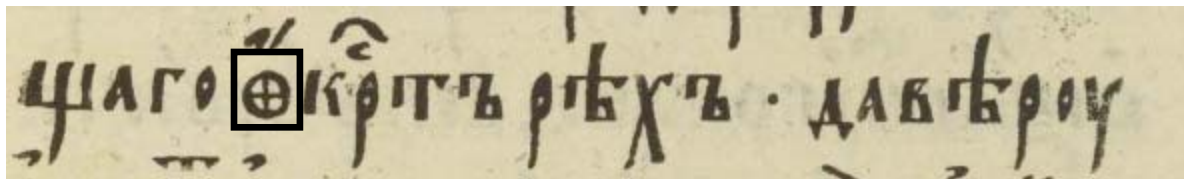






Figure 2: Extract from a 1553/4 Gospel showing the use of the Cyrillic crossed o (boxed).

Table 1: Summary of the Proposed Characters for Encoding

| Proposed Character | Proposed Name | Location | Comments |
|--|---|----------|---|
|  | CYRILLIC CAPITAL LETTER DOUBLE O | U + A698 | This letter is used in the words двое (two), обо (both), обанадесять (twelve), and двоюнадесять (twelve). |
|  | CYRILLIC SMALL LETTER DOUBLE O | U+A699 | |
|  | CYRILLIC CAPITAL LETTER CROSSED O | U + A69A | This letter is used in the word окрест (around, in the region of). |
|  | CYRILLIC SMALL LETTER CROSSED O | U+A69B | |

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

| | |
|--|---|
| 1. Title: | <i>Proposal to Encode Some Outstanding Early Cyrillic Characters in Unicode</i> |
| 2. Requester's name: | <i>Yuri Shardt, Nikita Simmons, Aleksandr Andreev</i> |
| 3. Requester type (Member body/Liaison/Individual contribution): | <i>Individual Contribution</i> |
| 4. Submission date: | <i>October 3, 2010</i> |
| 5. Requester's reference (if applicable): | |
| 6. Choose one of the following: | |
| This is a complete proposal: | <i>YES</i> |
| (or) More information will be provided later: | |

B. Technical – General

| | |
|---|--|
| 1. Choose one of the following: | |
| a. This proposal is for a new script (set of characters): | <i>NO</i> |
| Proposed name of script: | |
| b. The proposal is for addition of character(s) to an existing block: | <i>YES</i> |
| Name of the existing block: | <i>U+ A69x (Extended Cyrillic Block B)</i> |
| 2. Number of characters in proposal: | <i>4</i> |
| 3. Proposed category (select one from below - see section 2.2 of P&P document): | |
| A-Contemporary | <i>Yes</i> |
| B.1-Specialized (small collection) | |
| B.2-Specialized (large collection) | |
| C-Major extinct | |
| D-Attested extinct | |
| E-Minor extinct | |
| F-Archaic Hieroglyphic or Ideographic | |
| G-Obscure or questionable usage symbols | |
| 4. Is a repertoire including character names provided? | <i>YES</i> |
| a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? | <i>YES</i> |
| b. Are the character shapes attached in a legible form suitable for review? | <i>YES</i> |
| 5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? | <i>Yuri Shardt</i> |
| If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: | <i>Hirmos Ponomar v.6 (contact Yuri Shardt at yuri.shardt@ualberta.ca for the font)</i> |
| 6. References: | |
| a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? | <i>YES</i> |
| b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? | <i>YES</i> |
| 7. Special encoding issues: | |
| 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)? | <i>NO</i> |

8. Additional Information:

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 <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N3152-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05)

C. Technical - Justification

| | |
|--|---|
| 1. Has this proposal for addition of character(s) been submitted before? If YES explain | <i>NO</i> |
| 2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? If YES, available relevant documents: | <i>YES</i> <i>academics; Old Rite Believers</i> |
| 3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference: | <i>small, but active</i> |
| 4. The context of use for the proposed characters (type of use; common or rare) Reference: | <i>common</i> |
| 5. Are the proposed characters in current use by the user community? If YES, where? Reference: | <i>YES</i> <i>In typesetting older documents authentically</i> |
| 6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference: | <i>NO</i> |
| 7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? | <i>YES</i> |
| 8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference: | <i>NO</i> |
| 9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference: | <i>YES</i> <i>YES</i> |
| 10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference: | <i>NO</i> <i>NO</i> |
| 11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference: | <i>YES</i> <i>YES</i> |
| 12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary) | <i>NO</i> |
| 13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference: | <i>NO</i> |