PONOMAR PROJECT

Proposal to Encode the Mark's Chapter Glyph in the Unicode Standard

Aleksandr Andreev, Yuri Shardt, Nikita Simmons

1. <u>Introduction</u>

The symbols of the Russian Orthodox Typikon have already been proposed for inclusion in the Unicode standard (see (Shardt & Andreev, 2009) [n3772]). However, there remains one final glyph, the inclusion of which is necessary to properly typeset mediaeval and modern Slavonic liturgical texts within the framework of the Unicode Standard. This glyph is often referred to as "Mark's Chapter Glyph."

In Orthodox service books, Mark's Chapters are comments to difficult sections in the Typikon of the Lavra of St. Sabbas that were first compiled in the tenth century. These comments are attributed to a certain Monk Mark of the Lavra, possibly Bishop Mark of Hydruntum (Mansvetov, 1885, p. 219ff). These comments received a final revision when the Sabbaite Typikon was adopted by the Russian Orthodox Church in the fourteenth century. In their Russian version, they are indicated by a marginal glyph consisting of a stylized *Ⰳ* (Cyrillic Capital Em) and, often, other elements of the name Mark (in Cyrillic: *Ⰳарко*).

In the Russian Orthodox tradition, a total of three different forms can be found. The first form, which will be referred to as Type I, is shown in Figure 1 and Figure 2. This form dates from before the liturgical and orthographic reforms of Patriarch Nikon and is often seen in the *Oko Tserkovnoye* (*Typikon*) and the *Lenten Triodion* from this era. At present, Type I forms are occasionally still be used by Old Believers, who are those Orthodox that reject the reforms of Patriarch Nikon. It can be seen that Type I forms consistently shows various combinations of Cyrillic *Ⰳ, ρ,* and к. The last two letters can be located above or below the Cyrillic *Ⰳ*. Often, this glyph is in red type.

The second form, which will be referred to as Type II, is the Mark's Chapter Glyph that was standardised by reforms of Patriarch Nikon. A common representation of the Type II form is shown in Figure 3. It can be seen that the symbol now consists solely of *Ⰳ* and ρ and can occasionally be found in red.

Finally, there exist various variant forms, which will be referred to as Type III, for example, that shown in Figure 4, which encloses the complete name in a box.

Figure 1: Mark's Chapter Glyphs in the 1640 *Oko Tserkovnoye* published in Moscow.

Figure 2: Examples of Mark's Chapter Glyphs in the 1650 *Lenten Triodion* published in Moscow

Figure 3: Mark's Chapter Glyph in the 1986 *Typikon* published by the Moscow Patriarchate

Figure 4: Mark's Chapter Glyph in the 1893 *Menaion* published by the Kievan Lavra of the Caves

From the above figures, it is obvious that the form of the Mark's Chapter Glyph varies substantially between texts and time periods. While all glyphs have the Cyrillic letter capital Em, the other letters do not have fixed forms or positions. Although some of the Type I variants could be considered as containing combining (superscripted) Cyrillic letters ka and er, other Type I variants contain both superscripted and subscripted Cyrillic letters ka and er. On the other hand, Type II variants contain mostly a Cyrillic letter Capital Em with a combining Cyrillic letter er. Thus, it can be concluded that there many different forms for the Mark's Chapter Glyph.

It should be noted that while the form of Mark's Chapter Glyph varies substantially across texts, its function remains identical in all of the sources: to denote explanations given by

Monk Mark on difficult sections of the Typikon.

2. <u>Mark's Chapter Glyph in Existing Slavonic Standards</u>

The Ponomar Project (<u>http://www.ponomar.net/</u>) is pioneering the rendering, storage, and display of Slavonic-language liturgical texts in Unicode. Previous methods for encoding Church Slavonic include the Unified Church Slavonic (UCS), which uses the Windows-1251 codepage and assigns to it different values, and the Hyperinvariant Presentation (HIP) formats, which is a mark-up language that allows the required Slavonic characters to be entered using a Windows-1251 codepage.

In the UCS-8, which is the most recent version of UCS, there does not exist a unified approach to encoding the Mark's Chapter Glyph. This could be attributed to an oversight on part of the authors of this standard.

In the HIP format, the Mark's Chapter Glyph has been encoded as the unique command sequence <M\p> and distinguishes it clearly from the unique command M\p for a Cyrillic capital em with a superscripted Cyrillic er. In order to achieve backwards compatibility with this format, it would be advisable to include Mark's Chapter Glyph in Unicode.

3. <u>Existing Characters in Unicode</u>

Similar characters have already been encoded within the Unicode standard. However, their use is not a viable alternative. Perhaps the closest analogue is the Coptic Symbol Mi Ro (U+2CE5). However, the use of this symbol for the Mark's Chapter Glyph is not appropriate given the vastly distinct typographic and linguistic usages of the two characters[1].

4. <u>Justification for Inclusion of Mark's Chapter Glyph</u>

One approach to displaying the Mark's Chapter Glyph, especially in its Type II variant would be to use the Unicode sequence U+041C U+2DEC (Cyrillic capital em and combining Cyrillic er). However, this approach is problematic, as it would conflict with the ubiquitous abbreviation и꙼м꙼ка ("say name here"), which uses an м̂ without converting it into the Mark's Chapter Glyph form. It should be noted that both capital and lowercase versions of this

---

1   Not to mention that in academic contexts, one may also wish to include both Coptic and Slavonic texts.

abbreviation can be found. Finally, it can be noted that $\hat{м}$ of "say name here" and the Mark's Chapter Glyph have completely different functions and appearances. This implies that Mark's Chapter Glyph cannot be effectively rendered using a substitution table (such as for example GSUB in OpenType). This fact is noted by the HIPS standard which assigns different sequences of characters to represent the two entities. This strongly suggests that a separate codepoint should be created for Mark's Chapter Glyph.

As well, as the above figures show, this proposed sequence only describes Type II variants of this glyph. It does not reflect the forms found in older forms, especially the myriad Type I forms. In order to properly encode these forms, using composite glyphs, there would be a need to introduce subscript combining Cyrillic characters to the Unicode standard.

Finally, the glyph has an absolutely unique function and its inclusion in Unicode would greatly facilitate the storage, search, and editing of Slavonic-language liturgical texts.

5. **Summary**

In summary, the inclusion of the Mark's Chapter Glyph in the Miscellaneous Symbols block is proposed, following the previously proposed Typikon symbols. The proposed codepoint, representation, and name of the glyph are given in Table 1.

Table 1: Proposed Position and Representation of the Mark's Chapter Glypha

Proposed Codepoint	Representation	Proposed Name
U+1F545	ꙮ	Typikon Symbol Mark's Chapter

6. **Bibliography**

Mansvetov, I. (1885). *Церковный Устав (Типик) и его образование и судьба в Греческой и Русской Церкви. [The Church Ustav (Typikon) and its formation and usage in the Greek and Russian churches].* Moscow, Russian Empire.

Shardt, Y., & Andreev, A. (2009). *Proposal to Encode the Typikon Symbols in Unicode (L2/09-310).* Proposal.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
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:** *Proposal to Encode the Mark's Chapter Glyph in the Unicode Standard*
2. Requester's name: *Aleksandr Andreev, Yuri Shardt, Nikita Simmons,*
3. Requester type (Member body/Liaison/Individual contribution): *Individual Contribution*
4. Submission date: *October 3, 2010*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *YES*
   (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): *NO*
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *YES*
      Name of the existing block: *U+ 1F54x*
2. Number of characters in proposal: *1*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary ___ B.1-Specialized (small collection) *Yes* 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? *YES*
      a. If YES, are the names in accordance with the "character naming guidelines"
         in Annex L of P&P document? *YES*
      b. Are the character shapes attached in a legible form suitable for review? *YES*
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? *Yuri Shardt*
      If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: *Hirmos Ponomar v.6 (contact Yuri Shardt at yuri.shardt@ualberta.ca for the font)*
6. References:
      a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *YES*
      b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *YES*
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)? *NO*

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.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?   *NO*
    If YES explain
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.)?   *YES*
        If YES, with whom?   *academics; publishers of service books*
        If YES, available relevant documents:
3. Information on the user community for the proposed characters (for example:
    size, demographics, information technology use, or publishing use) is included?   *>300 million*
    Reference:
4. The context of use for the proposed characters (type of use; common or rare)   *common*
    Reference:
5. Are the proposed characters in current use by the user community?   *YES*
    If YES, where? Reference: *In typesetting religious service books*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
    in the BMP?   *NO*
        If YES, is a rationale provided?
        If YES, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?   *YES*
8. Can any of the proposed characters be considered a presentation form of an existing
    character or character sequence?   *NO*
        If YES, is a rationale for its inclusion provided?
        If YES, reference:
9. Can any of the proposed characters be encoded using a composed character sequence of either
    existing characters or other proposed characters?   *YES*
        If YES, is a rationale for its inclusion provided?   *YES*
        If YES, reference:
10. Can any of the proposed character(s) be considered to be similar (in appearance or function)
    to an existing character?   *NO*
        If YES, is a rationale for its inclusion provided?   *NO*
        If YES, reference:
11. Does the proposal include use of combining characters and/or use of composite sequences?   *YES*
    If YES, is a rationale for such use provided?   *YES*
        If YES, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
        If YES, reference:
12. Does the proposal contain characters with any special properties such as
    control function or similar semantics?   *NO*
        If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility character(s)?   *NO*
    If YES, is the equivalent corresponding unified ideographic character(s) identified?
        If YES, reference: