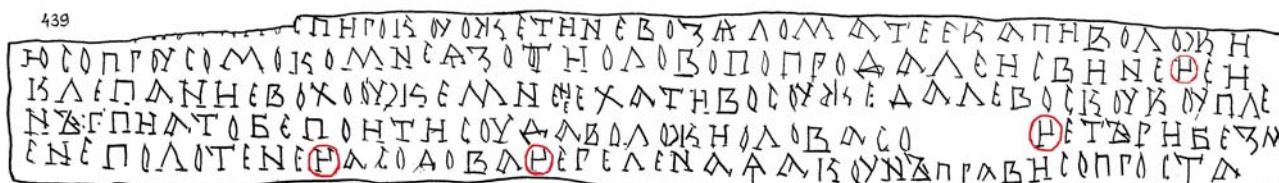Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation
Международная организация по стандартизации

Doc Type: **Working Group Document**
Title: **Proposal to encode two Cyrillic characters in the BMP of the UCS**
Source: **Ralph Cleminson and Michael Everson**
Status: **Individual Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2009-01-26**

1. Character additions. This document requests the addition of two new Cyrillic characters to the UCS. They are used in certain early Russian documents and are necessary if these are to be adequately encoded. For conformity with the rest of the block, both a capital and a small character are requested.

Ц ц

REVERSED TSE Ц ц (in Slavonic style **Ц ц**) is commonly found in documents produced in Novgorod and its environs from the 11th to the 15th centuries. In the language of this area the distinction between ц (ts) and ч (ʧ) had been eliminated, and ц replaces both these characters in the documents. It cannot be considered equivalent to either of them, and therefore neither can replace it in transcription.

Figure 1. Novgorod Birch-Bark Letter № 439 (beginning of the 13th century). This shows four examples of ц, of which the first and third correspond to etymological ц and the second and fourth to ч. The characters ц and ч are not used. Published in: А. В. Арциховский, В. Л. Янин, *Новгородские грамоты на бересте (из раскопок 1962-1976 гг.)*. Москва: Наука, 1978, pp.41-45.

2. Ordering. The relative position of the CYRILLIC REVERSED TSE is shown in red below.

3. Unicode Character Properties. Character properties are proposed here.

```
A660;CYRILLIC CAPITAL LETTER REVERSED TSE;Lu;0;L;;;;;N;;;;A661;
A661;CYRILLIC SMALL LETTER REVERSED TSE;Ll;0;L;;;;;N;;;A660;;A660
```

4. Bibliography.

Живов В. М. 2006. *Восточнославянское правописание XI–XIII века*. РАН; Ин-т русского языка им. В.В.Виноградова. М.: Языки славянской культуры, 2006. С. 100, ссылка 22

Попов, М. Б. 2005. *Проблемы синхронической и диахронической фонологии русского языка*. СПб. С. 214215

Worth Dean S. 1985. "Mirror Reversals in Novgorod Paleography", in *Language and Literary Theory: (In Honor of Ladislav Matejka)*. (Papers of Slavic Philology; № 5)Ann Arbor: University of Michigan. Pp. 215-222.

Row A6: CYRILLIC EXTENDED-B

	A64	A65	A66	A67	A88	A89
0	Ꙁ	Ꙋ	Ꙡ	⸀	Ꚁ	Ꚁ
1	ꙁ	ꙋ	ꙡ	⸁	ꚁ	ꚁ
2	Ꙃ	Ꙍ	Ꙣ	⸂	ꙷ	Ꚃ
3	ꙃ	ꙍ	Ꙣ	⸃	ꙷ	ꚃ
4	Ꙅ	Ꙏ	Ꙣ		ЖК	Ꚅ
5	ꙅ	ꙏ	ꙣ		жк	ꚅ
6	Ꙇ	Ꙑ	Ꙥ		Ꚇ	Ꚉ
7	ꙇ	ꙑ	ꙥ		ꚇ	ꚉ
8	Ꙉ	Ꙓ	Ꙧ		Ꚁ	
9	ꙉ	ꙓ	ꙧ		ꚁ	
A	Ꙋ	Ꙕ	Ꙩ		Ꙋ	
B	ꙋ	ꙕ	ꙩ		ꙋ	
C	Ꙍ	Ꙗ	Ꙫ	꙼	Ꚍ	
D	ꙍ	ꙗ	ꙫ	꙽	ꚍ	
E	Ꙏ	Ꙙ	Ꙭ	꙾	Ꚏ	
F	ꙏ	ꙙ	ꙭ	ꙿ	ꚏ	

Row A6: CYRILLIC EXTENDED-B

hex	Name
40	CYRILLIC CAPITAL LETTER ZEMLYA
41	CYRILLIC SMALL LETTER ZEMLYA
42	CYRILLIC CAPITAL LETTER DZELO
43	CYRILLIC SMALL LETTER DZELO
44	CYRILLIC CAPITAL LETTER REVERSED DZE
45	CYRILLIC SMALL LETTER REVERSED DZE
46	CYRILLIC CAPITAL LETTER IOTA
47	CYRILLIC SMALL LETTER IOTA
48	CYRILLIC CAPITAL LETTER DJERV
49	CYRILLIC SMALL LETTER DJERV
4A	CYRILLIC CAPITAL LETTER MONOGRAPH UK
4B	CYRILLIC SMALL LETTER MONOGRAPH UK
4C	CYRILLIC CAPITAL LETTER BROAD OMEGA
4D	CYRILLIC SMALL LETTER BROAD OMEGA
4E	CYRILLIC CAPITAL LETTER NEUTRAL YER
4F	CYRILLIC SMALL LETTER NEUTRAL YER
50	CYRILLIC CAPITAL LETTER YERU WITH BACK YER
51	CYRILLIC SMALL LETTER YERU WITH BACK YER
52	CYRILLIC CAPITAL LETTER IOTIFIED YAT
53	CYRILLIC SMALL LETTER IOTIFIED YAT
54	CYRILLIC CAPITAL LETTER REVERSED YU
55	CYRILLIC SMALL LETTER REVERSED YU
56	CYRILLIC CAPITAL LETTER IOTIFIED A
57	CYRILLIC SMALL LETTER IOTIFIED A
58	CYRILLIC CAPITAL LETTER CLOSED LITTLE YUS
59	CYRILLIC SMALL LETTER CLOSED LITTLE YUS
5A	CYRILLIC CAPITAL LETTER BLENDED YUS
5B	CYRILLIC SMALL LETTER BLENDED YUS
5C	CYRILLIC CAPITAL LETTER IOTIFIED CLOSED LITTLE YUS
5D	CYRILLIC SMALL LETTER IOTIFIED CLOSED LITTLE YUS
5E	CYRILLIC CAPITAL LETTER YN
5F	CYRILLIC SMALL LETTER YN
60	CYRILLIC CAPITAL LETTER REVERSED TSE
61	CYRILLIC SMALL LETTER REVERSED TSE
62	CYRILLIC CAPITAL LETTER SOFT DE
63	CYRILLIC SMALL LETTER SOFT DE
64	CYRILLIC CAPITAL LETTER SOFT EL
65	CYRILLIC SMALL LETTER SOFT EL
66	CYRILLIC CAPITAL LETTER SOFT EM
67	CYRILLIC SMALL LETTER SOFT EM
68	CYRILLIC CAPITAL LETTER MONOCULAR O
69	CYRILLIC SMALL LETTER MONOCULAR O
6A	CYRILLIC CAPITAL LETTER BINOCULAR O
6B	CYRILLIC SMALL LETTER BINOCULAR O
6C	CYRILLIC CAPITAL LETTER DOUBLE MONOCULAR O
6D	CYRILLIC SMALL LETTER DOUBLE MONOCULAR O
6E	CYRILLIC LETTER MULTIOCULAR O
6F	COMBINING CYRILLIC VZMET
70	COMBINING CYRILLIC TEN MILLIONS SIGN
71	COMBINING CYRILLIC HUNDRED MILLIONS SIGN
72	COMBINING CYRILLIC THOUSAND MILLIONS SIGN
73	SLAVONIC ASTERISK
74	(This position shall not be used)
75	(This position shall not be used)
76	(This position shall not be used)
77	(This position shall not be used)
78	(This position shall not be used)
79	(This position shall not be used)
7A	(This position shall not be used)
7B	(This position shall not be used)
7C	COMBINING CYRILLIC KAVYKA
7D	COMBINING CYRILLIC PAYEROK
7E	CYRILLIC KAVYKA
7F	CYRILLIC PAYEROK
80	CYRILLIC CAPITAL LETTER DWE
81	CYRILLIC SMALL LETTER DWE
82	CYRILLIC CAPITAL LETTER DZWE
83	CYRILLIC SMALL LETTER DZWE
84	CYRILLIC CAPITAL LETTER ZHWE
85	CYRILLIC SMALL LETTER ZHWE
86	CYRILLIC CAPITAL LETTER CCHE
87	CYRILLIC SMALL LETTER CCHE
88	CYRILLIC CAPITAL LETTER DZZE
89	CYRILLIC SMALL LETTER DZZE
8A	CYRILLIC CAPITAL LETTER TE WITH MIDDLE HOOK
8B	CYRILLIC SMALL LETTER TE WITH MIDDLE HOOK
8C	CYRILLIC CAPITAL LETTER TWE
8D	CYRILLIC SMALL LETTER TWE
8E	CYRILLIC CAPITAL LETTER TSWE
8F	CYRILLIC SMALL LETTER TSWE
90	CYRILLIC CAPITAL LETTER TSSE
91	CYRILLIC SMALL LETTER TSSE
92	CYRILLIC CAPITAL LETTER TCHE
93	CYRILLIC SMALL LETTER TCHE
94	CYRILLIC CAPITAL LETTER HWE
95	CYRILLIC SMALL LETTER HWE
96	CYRILLIC CAPITAL LETTER SHWE
97	CYRILLIC SMALL LETTER SHWE
98	(This position shall not be used)

hex	Name
99	(This position shall not be used)
9A	(This position shall not be used)
9B	(This position shall not be used)
9C	(This position shall not be used)
9D	(This position shall not be used)
9E	(This position shall not be used)
9F	(This position shall not be used)

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 add Cyrillic Reversed Tse*
2. Requester's name: *Professor R.M.Cleminson & Michael Everson*
3. Requester type (Member body/Liaison/Individual contribution): *individual contribution*
4. Submission date: *2009.01.20*
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):
         Proposed name of script:
     b. The proposal is for addition of character(s) to an existing block: *yes*
         Name of the existing block: *Cyrillic Extended - B*
2. Number of characters in proposal:
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary   B.1-Specialized (small collection)   x   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? *The proposers*
     If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools
     used: *http://www.obshtezhitie.net/res/dilyan.ttf*
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? *Discussed within the International Committee of Slavists'* Commission on
   the Computer-Supported Processing of Mediæval Slavonic
   Manuscripts and Early Printed 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?
   Reference: *Historians, linguists, archaeologists*

4. The context of use for the proposed characters (type of use; common or rare)
   Reference: *Necessary for scholarly publications of mediaeval documents*

5. Are the proposed characters in current use by the user community? yes
   If YES, where? Reference: *see below*

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
   in the BMP? *yes*
   If YES, is a rationale provided? *yes*
   If YES, reference: *The Cyrillic alphabet belongs within the BMP.*

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? *no*
   If YES, is a rationale for its inclusion provided?
   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?
    If YES, reference:

11. Does the proposal include use of combining characters and/or use of composite sequences? *no*
    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:

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: