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 four Cyrillic characters in the BMP of the UCS**
Source: **Michael Everson and Soslan Khubulov**
Status: **Individual Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2012-01-31**

1. Character additions. This document requests the addition of four new Cyrillic characters to the UCS. Two of these letters were introduced into the Ossetian alphabet in 1844 by Andreas Johan Sjögren. That alphabet was in use up to 1924 when the Latin script was adopted for Ossetian. There is a great deal of literature written during the 80 years when that orthography was in use. Both a capital and a small character are requested.

Those same two letters and two more are attested in a Komi grammar from 1850. Thus the following four characters are requested:

052A	ДҜ	CYRILLIC CAPITAL LETTER DZZHE
052B	дҝ	CYRILLIC SMALL LETTER DZZHE
		• Ossetian, Komi
052C	Ӌ	CYRILLIC CAPITAL LETTER DCHE
052D	ӌ	CYRILLIC SMALL LETTER DCHE
		• Komi

2. Ordering. The relative position of these two letters (taking case pairing as read) is shown below.

DZZHE follows immediately after ZHE and before ZHWE
DCHE follows immediately after CHE and before TCHE

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

```
052A;CYRILLIC CAPITAL LETTER DZZHE;Lu;0;L;;;;;N;;;;052B;
052B;CYRILLIC SMALL LETTER DZZHE;Ll;0;L;;;;;N;;;052A;;052A
052C;CYRILLIC CAPITAL LETTER DCHE;Lu;0;L;;;;;N;;;;052D;
052D;CYRILLIC SMALL LETTER DCHE;Ll;0;L;;;;;N;;;052C;;052C
```
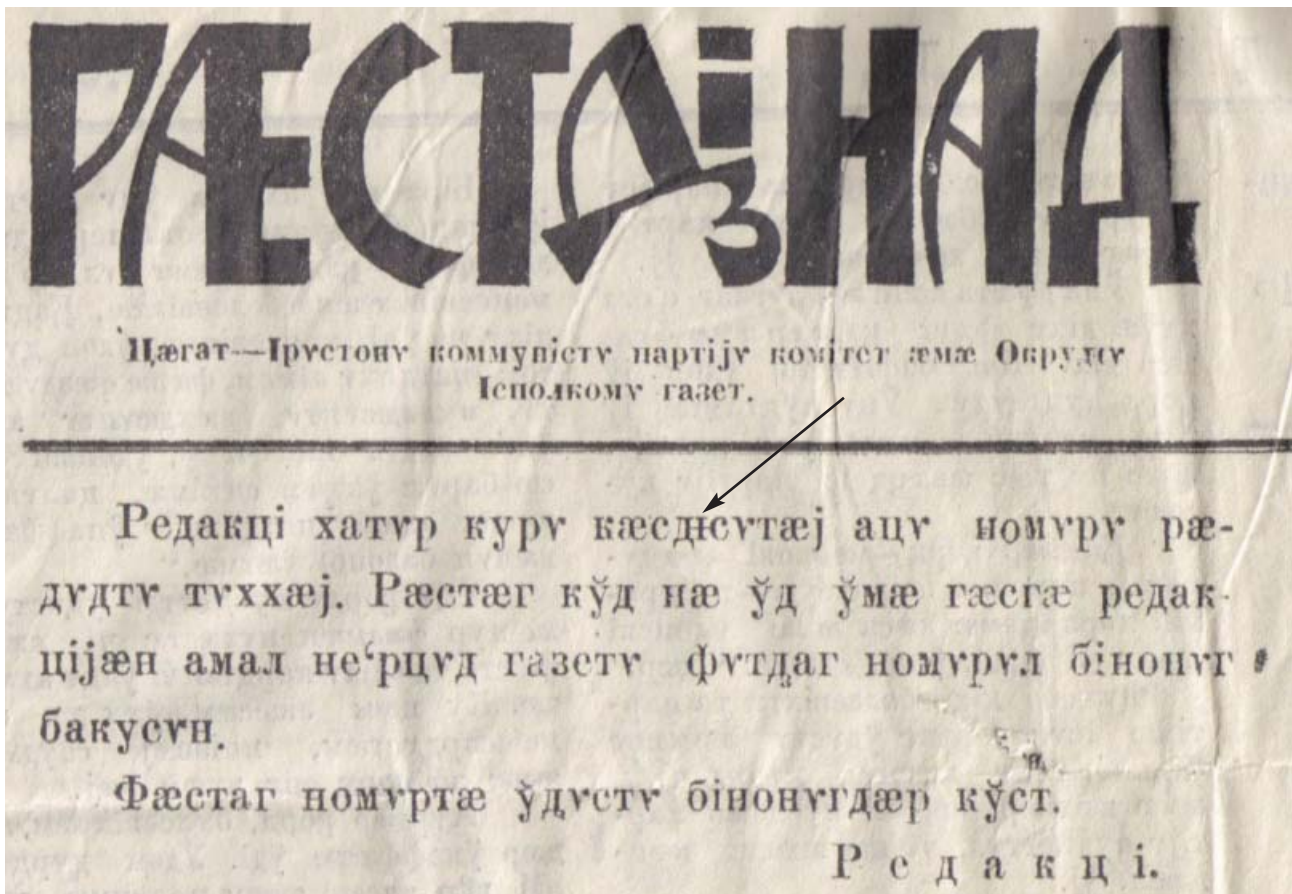
Figures

Figure 1. Sample from the gazette *Ræstₓinad* showing the word *kæsₓ̂ỹtæj* using the letter DZHE.

Figure 2. Sample showing the letter DZZHE.
Of interest to typographers is the italic form of дж DZZHE, д DZZE, and тц TSSE shown here.

9. Кувгæ та кæнут афтæ: Мах Фvд, кæцv дæ æрвтv мiдæг! свӻдæг ўæд Дæ ном:

10. Æрцæўæд Дæ патцахад; Дæ бар ўæд арбv кўд у, зæххvл дæр афтæ;

11. Нæ дзул нvн онvнæн радт махæн абон;

12. Æмæ нvн нvппар нæ хæстæ, мах нæ хæс-дӀсvнтæн кўд барæм, афтæ;

13. Æмæ нæ ма бакæн фvдæвзарæнv; фæлæ нæ фервæзvнкæн фvдбvлvзæj. Цæмæjдæрiдтæр Дæў у патцахад, æмæ тvх, æмæ кад мvккагмæ. Амiн.

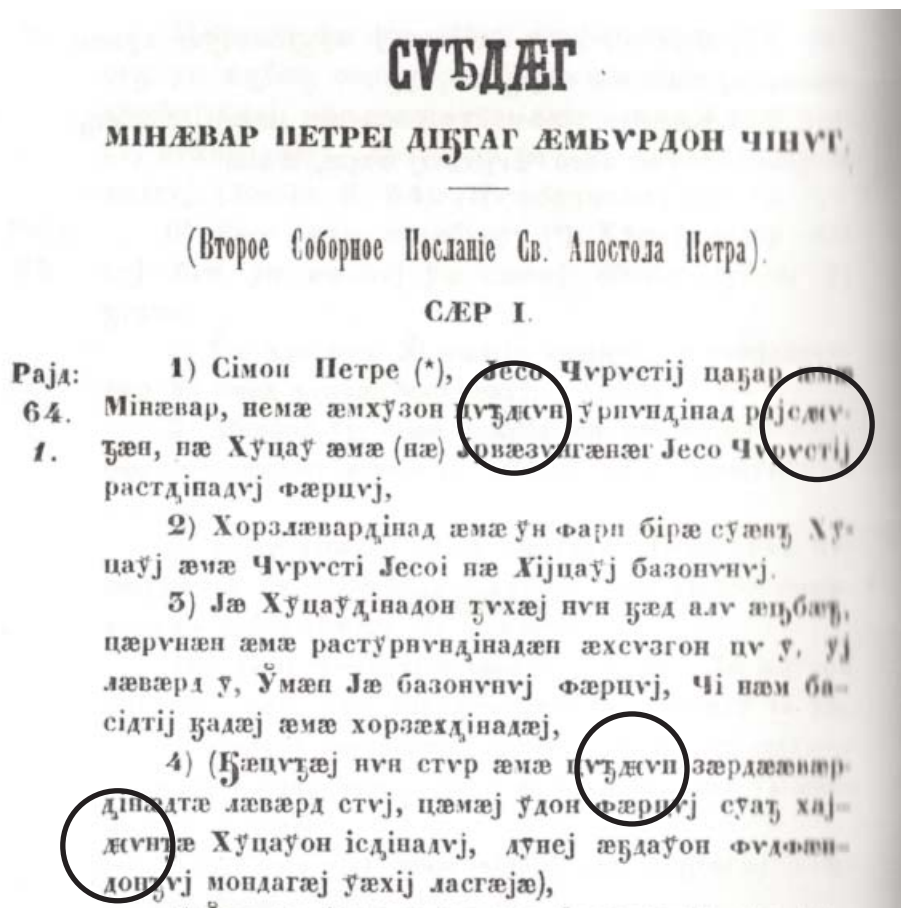**Figure 3.** Sample religious text showing the letter ᴅᴢᴢʜᴇ in the word *xæsžùntæn*.

Figure 4. Sample showing numerous examples of the letter ᴅᴢᴢʜᴇ.

да, въ Зырянскомъ языкѣ находится *тридцать буквъ,*
именно:

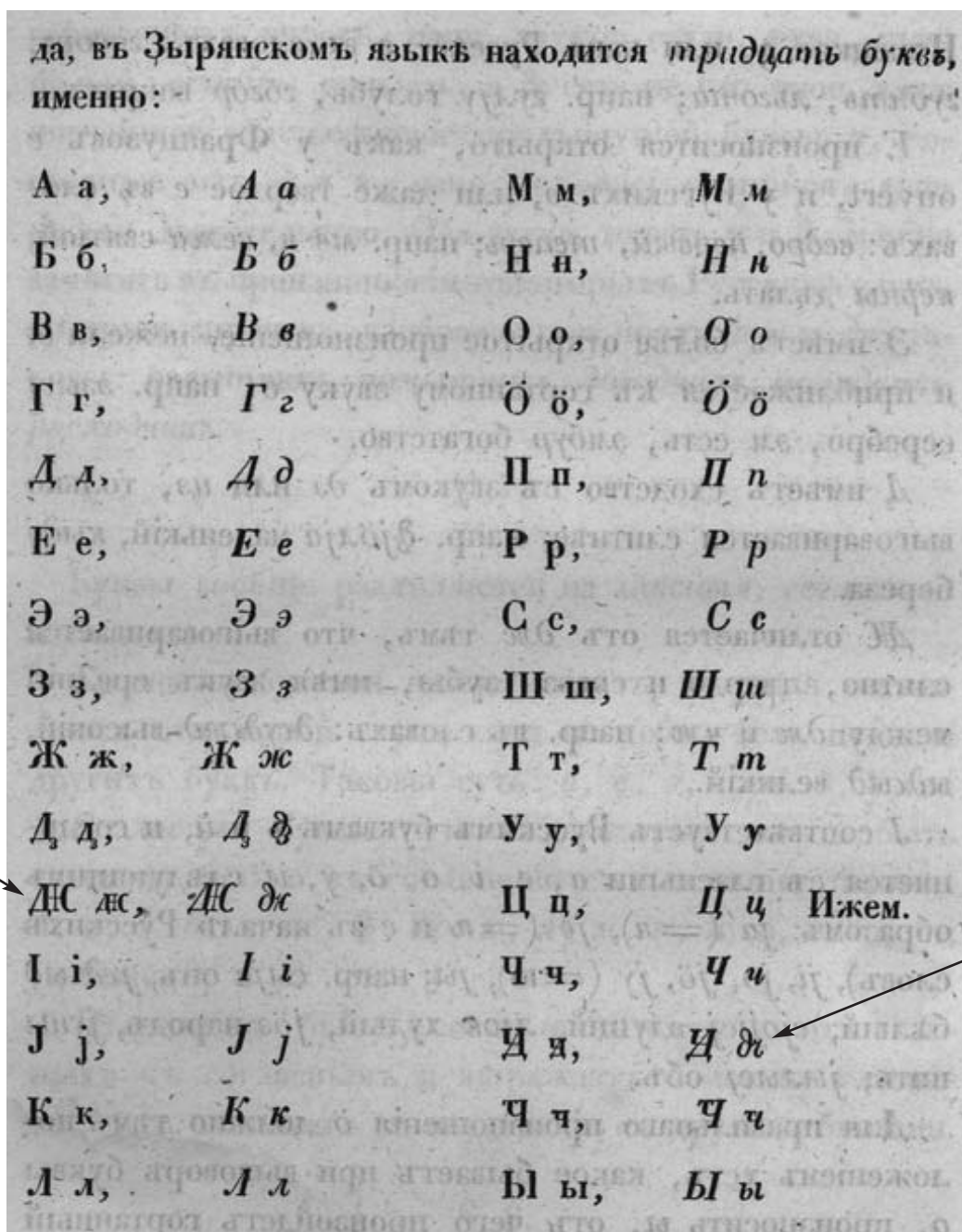|			
А а,	*А а*	М м,	*М м*
Б б,	*Б б*	Н н,	*Н н*
В в,	*В в*	О о,	*О о*
Г г,	*Г г*	Ö ö,	*Ö ö*
Д д,	*Д д*	П п,	*П п*
Е е,	*Е е*	Р р,	*Р р*
Э э,	*Э э*	С с,	*С с*
З з,	*З з*	Ш ш,	*Ш ш*
Ж ж,	*Ж ж*	Т т,	*Т т*
Д̆ д̆,	*Д̆ д̆*	У у,	*У у*
ДС ӕс,	*ДС дж*	Ц ц,	*Ц ц* Ижем.
I i,	*I i*	Ч ч,	*Ч ч*
J j,	*J j*	Д̌ я,	*Д̌ дж*
К к,	*К к*	Ч ч,	*Ч ч*
Л л,	*Л л*	Ы ы,	*Ы ы*

Figure 5. Sample showing the letters DZZHE and DCHE as well as the the letter DZZE, from Pavel Savvaitov's *Grammatika Zyrjanskago jazyka* (Sanktpeterburg 1850)

ДС отличается отъ *дж* тѣмъ, что выговаривается слитно, твердо и сквозь зубы, имѣя звукъ средній между *дж* и *чж*; напр. въ словахъ: *джуджыд* высокій, *ыджыд* великій.

Figure 6. Sample describing the phonetics of theh letter DZZHE.

Ч есть усиленное ч, состоящее изъ *т* и *ч*, выго-
вариваемыхъ вмѣстѣ, напр. *чjукöр* стадо, куча, *iчет*
малый, *вiчiсjны* ожидать; а *дч* есть не что иное, какъ
мягчайшее видоизмѣненіе предъидущей буквы ч, со-
стоящее изъ *д* и *ч*, напр. *велöдiсj* учащійся, *лjас-
кöдчем* ласкательство. Оба звука, то есть ч и *дч*, можно
замѣтить въ произношеніи нѣкоторыхъ Русскихъ словъ,
которыя впрочемъ изображаются чрезъ *тч* и *дч*; та-
ковы: *повытчикъ, помытчикъ, доводчикъ, молодчикъ,
расходчикъ.*

Figure 7. Sample describing the phonetics of the letters TCHE and DCHE.

A. Administrative

1. Title

Proposal to encode four Cyrillic letters in the BMP of the UCS

2. Requester's name

Michael Everson and Soslan Khubulov

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

4. Submission date

2012-01-30

5. Requester's reference (if applicable)

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)

No.

1b. Proposed name of script

1c. The proposal is for addition of character(s) to an existing block

Yes.

1d. Name of the existing block

Cyrillic Supplement.

2. Number of characters in proposal

2.

3. Proposed category (A-Contemporary; 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)

Category A.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson. Fontlab.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

No.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. 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.

8. 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.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

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?

Soslan Khubulov is a user.

2c. 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?

Historians of Komi.

4a. The context of use for the proposed characters (type of use; common or rare)
Rare today; common previously.
4b. Reference
5a. Are the proposed characters in current use by the user community?
Yes.
5b. If YES, where?
Russian Federation.
6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?
Yes.
6b. If YES, is a rationale provided?
Yes.
6c. If YES, reference
Keep with other Cyrillic letters.
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
No.
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?
No.
11b. If YES, is a rationale for such use provided?
11c. If YES, reference
11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
No.
11e. If YES, reference
12a. Does the proposal contain characters with any special properties such as control function or similar semantics?
No.
12b. If YES, describe in detail (include attachment if necessary)
13a. Does the proposal contain any Ideographic compatibility character(s)?
No.
13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?