Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международнаяорганизацияпостандартизации

DocType: WorkingGroupDocument
Title: Proposal to encode Arabic characters used for Adyghe and Chechen languages
Source: See References
Authors: Ilya Yevlampiev, Karl Pentzlin, Nurlan Joomagueldinov
Status: Expert Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 19 July 2011

1. Introduction

The letters of Adyghe and Chechenold Arabic orthographies that are not encoded in Unicode standard are presented here.

2. Proposed Characters

Annotations below address special issues for a character, or reference to figures where such special issues are discussed. (These annotations are not intended to be retained in the character list when copied into the standard.)

Proposed codepoint	Glyph	F M I	Proposed name	Linguistic comment
08B2	ﺟ	ﺟﺠﺠ	ARABIC LETTER HAH WITH SMALL V BELOW • Adyghe	*now is denoted by Ч*
08B3	ﺟ	ﺟﺠﺠ	ARABIC LETTER HAH WITH INVERTED DAMMA BELOW • Adyghe	*now is denoted by КІ*
08B4	ﻳ	ﻳ··	ARABIC LETTER REH WITH TWO DOTS BELOW • Chechen	*now is denoted by ЦІ*
08B5	ﻳ	ﻳ··	ARABIC LETTER REH WITH THREE DOTS POINTING UPWARDS BELOW • Chechen	*now is denoted by Ц*
08B6	ﺵ	ﺵﺘﺶ	ARABIC LETTER SEEN WITH TWO DOTS ABOVE • Adyghe	*now is denoted by Ш*
08B7	ﻙ	ﻙﻜﻚ	ARABIC LETTER ARABIC LETTER KEHEH WITH HALF BAR • Adyghe	*now is denoted by КІ*
08B8	ﻝ	ﻝﻠﻞ	ARABIC LETTER LAM WITH HALF BAR • Adyghe	*now is denoted by ЛІ*
08B9	ﻝ	ﻝﻠﻞ	ARABIC LETTER LAM WITH DOUBLE HALF BAR • Adyghe	*now is denoted by ЛЪ*

3. Encoding Considerations

```
Joining type and group for ArabicShaping.txt:
08B2; HAH WITH SMALL V BELOW; D; HAH
08B3; HAH WITH INVERTED DAMMA BELOW; D; HAH
08B4; REH WITH 2 DOTS BELOW; R; REH
08B5; REH WITH 3 DOTS POINTING UPWARDS BELOW; R; REH
08B6; SEEN WITH 2 DOTS ABOVE; D; SEEN
08B7; KEHEH WITH HALF BAR; D; GAF
08B8; LAM WITH HALF BAR; D; LAM
08B9; LAM WITH 2 HALF BAR; D; LAM


Unicode character Properties:
08B2;ARABIC LETTER HAH WITH SMALL V BELOW;Lo;0;AL;;;;;N;;;;;
08B3;ARABIC LETTER HAH WITH INVERTED DAMMA
BELOW;Lo;0;AL;;;;;N;;;;;
08B4;ARABIC LETTER REH WITH TWO DOTS BELOW;Lo;0;AL;;;;;N;;;;;
08B5;ARABIC LETTER REH WITH THREE DOTS POINTING UPWARDS
BELOW;Lo;0;AL;;;;;N;;;;;
08B6;ARABIC LETTER SEEN WITH TWO DOTS ABOVE;Lo;0;AL;;;;;N;;;;;
08B7;ARABIC LETTER KEHEH WITH HALF BAR;Lo;0;AL;;;;;N;;;;;
08B8;ARABIC LETTER LAM WITH HALF BAR;Lo;0;AL;;;;;N;;;;;
08B9;ARABIC LETTER LAM WITH DOUBLE HALF BAR;Lo;0;AL;;;;;N;;;;;


Collation:
```

Characters may be sorted after similar characters.
- 08B2 ARABIC LETTER HAH WITH SMALL V BELOW and 08B3 ARABIC LETTER HAH WITH INVERTED DAMMA BELOW should sort after U+08A2 ARABIC LETTER JEEM WITH TWO DOTS ABOVE
- 08B4 ARABIC LETTER REH WITH TWO DOTS BELOW and 08B5 ARABIC LETTER REH WITH THREE DOTS POINTING UPWARDS BELOW should sort after U+0771 ARABIC LETTER REH WITH SMALL ARABIC LETTER TAH AND TWO DOTS and before 08AA ARABIC LETTER REH WITH LOOP
- 08B6 ARABIC LETTER SEEN WITH TWO DOTS ABOVE should sort after U+077E ARABIC LETTER SEEN WITH INVERTED V
- 08B7 ARABIC LETTER ARABIC LETTER KEHEH WITH HALF BAR should sort after U+0764 ARABIC LETTER KEHEH WITH THREE DOTS POINTING UPWARDS BELOW
- 08B8 ARABIC LETTER LAM WITH HALF BAR and 08B9 ARABIC LETTER LAM WITH DOUBLE HALF BAR should sort after U+08A6 ARABIC LETTER LAM WITH DOUBLE BAR

Confusability

Confusability issues would only arise if any of these characters were decomposed.

4. Usage and Encoding Rationale

`Additions for Adyghe`

The proposed letter was used in unified Adyghe (Agygean, Kabardian, Circassian) alphabet in 1918-27. The example is found there:

http://ru.wikipedia.org/wiki/%D0%A4%D0%B0%D0%B9%D0%BB:Adyghe_arabic_alphabet.JPG

Contextual shaping for ARABIC LETTER LAM WITH HALF BAR and ARABIC LETTER LAM WITH DOUBLE HALF BAR is similar to that of U+0644 ARABIC LETTER LAM and U+076A ARABIC LETTER LAM WITH BAR, including the requirement to form ligatures with ALEF and related characters. For examples see fig. 1, 5, 6.

`Additions for Chechen`

The first proposed letter was used in 1922-1925 orthography of Chechen language. This orthography was based on the Persian and Turkic writing tradition. See an example of such writing here:

http://ru.wikipedia.org/wiki/%D0%A4%D0%B0%D0%B9%D0%BB:Chechen_alphabet-1925.gif.

and here

http://fotki.yandex.ru/users/ievlampiev/album/66437/?&p=1

The second proposed letter was found in diasporas' Arabic orthography, the scan image provided by Noorlan Joomagueldinov (see examples section). The orthography is mainly used in religious literature and is still alive. This orthography is based on pure Arabic writing tradition, without usage of Persian and Turkic Arabic letters. For examples see fig. 2, 3, 4.
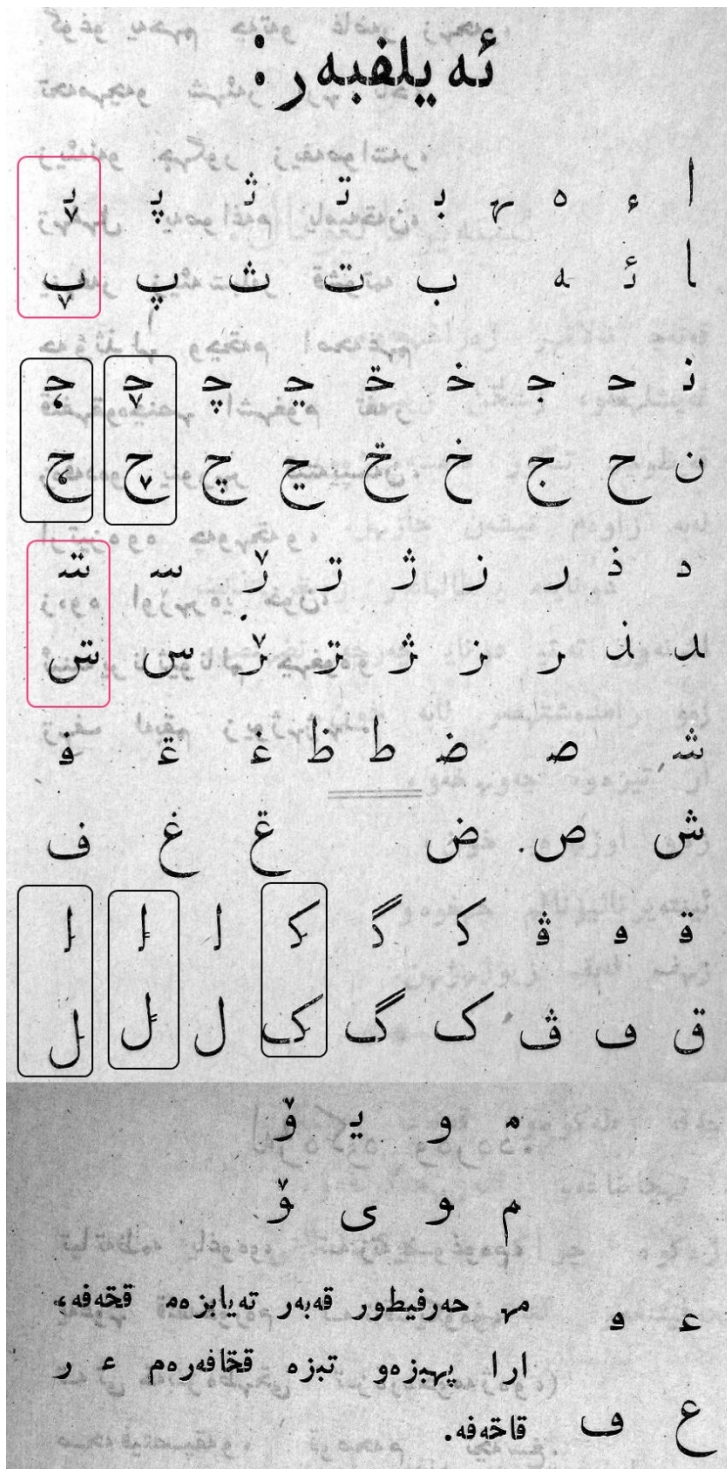
5. Acknowledgements

Special thanks to Lorna Priest for valuable help in composing this proposal.

6. References

[1] Хатанов А., Хатков А. Учись-трудись. Сельскохозяйственный букварь для взрослых. М., 1924

[2] Чеченская азбука. Буйнакск, 1925

[3] Литературная энциклопедия: В 11 т. — [М.], 1929—1939.

[4] Н. В. Юшманов. Определитель языков. М.-Л., 1941

7. Examples and Figures

Fig. 1. Adyghe alphabet table. The proposed letters are in frames. For running text examples some reverse shadows are encircled in the second image [1].

ܡܒܦܠܝܡܐ

Fig. 2.Chechen alphabet of 1922-1925, unsorted table. The proposed letter is in frame [2].

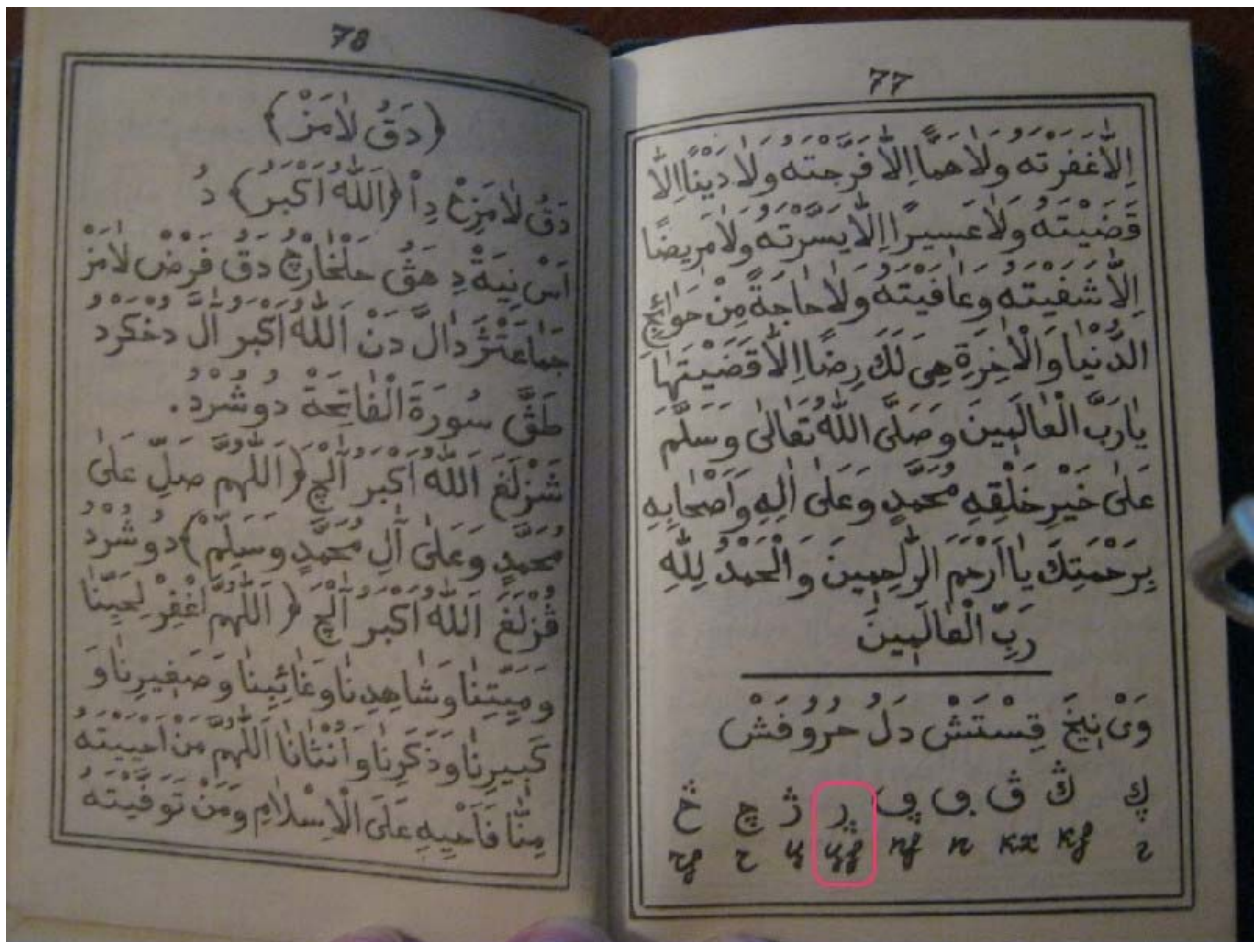Fig. 3.The primer page devoted to the proposed letter [2].

Fig. 4.The religious Veinakh book printed in Turkey (from Noorlan Joomagueldinov's collection), supplied with transliteration to the modern Cyrillic alphabet. The proposed letter is into the frame. In the second image the proposed letter is shown in the running text.

(مسح در)

مسج دش ولچن محبسش لامز كرحا
دلش طيخن خل بيز . نارع ولج
دينج ببى مسج دملك . نيفز أرنث
ولج قاد ينج قابسى مسج دملك . اخن
چق يلج مسج دوغ . كوكر محبى دع
يقچا مسج دوغ . كوكر محبى دع يقن
دخن لامز رش كك بلج كرا خل . يخن
يطا محيشثر مسج رمكد .

وسلم على خير خلقه محمد واله وصحبه
اجمعين

(تيزيتح)

الحمد لله رب العالمين والصلاة والسلام
على رسول الله اللهم اغفرله ها وارحمه ها
واجعله جا من أهل الجنة واجره جا من عذاب
القبر والنار اللهم صل على محمد وعلى
ال محمد وسلم بحرمة سورة الفاتحة

ياآله العالمين وياخير النا صرين
توفنامسلمين والحقنا بالصالحين

Fig. 5. The correspondence table between the Latin and Arabic Adyghe alphabet from the Soviet Literature Encyclopaedia 1929-39 [3]. This is one of the most competitive sources on the early Soviet writing systems. Proposed letters are encircled.

Fig. 6. Adyghe in Yushmanov's Определитель языков (Determinant of Languages) [4]. Only the most frequent and distinctive letters are shown. The proposed letters are encircled.
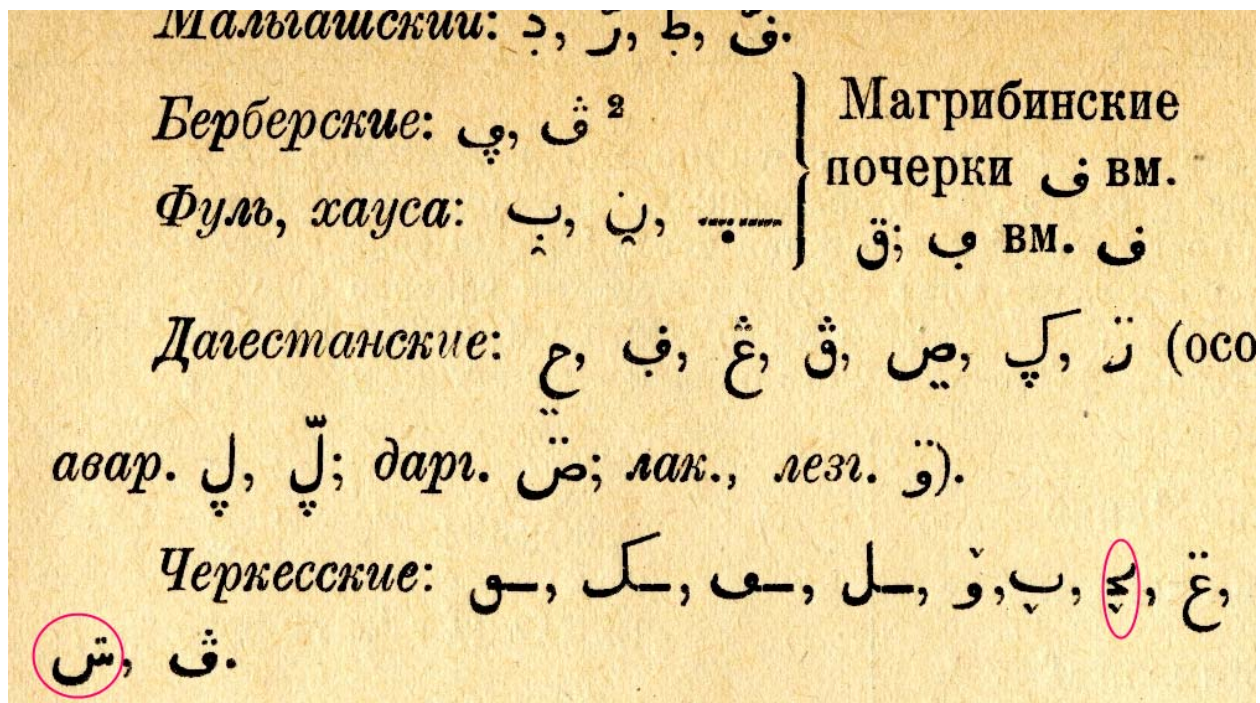
ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARYFORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and 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:** *Revised proposal to encode Arabic characters used for Adyghe and Chechen languages*
2. Requester's name: *Ilya Yevlampiev, Karl Pentzlin, Nurlan Joomagueldinov*
3. Requester type (Member body/Liaison/Individual contribution): *Expert Contribution*
4. Submission date: *19 July 2011*
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: *No*

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: *Arabic Extended-A*
2. Number of characters in proposal: *8*
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. Fonts related:
        a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
    Lorna Priest, SIL International
        b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):

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)? *Yes*
    Sorting and linguistic representations are discussed in the proposal

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] Form number: N3702-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, 2009-11)

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? *Linguists, librarians*
        If YES, available relevant documents: *See examples in proposal*
3. Information on the user community for the proposed characters (for example:
    size, demographics, information technology use, or publishing use) is included? *once used by 1 million people*
    Reference:
4. The context of use for the proposed characters (type of use; common or rare) *Historically common*
    Reference:
5. Are the proposed characters in current use by the user community? Historical use
    If YES, where? Reference:
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? *Should be placed with similar characters*
        If YES, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
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? *Discussion in proposal*
        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: