

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 add four characters for Kashmiri to the BMP of the UCS****Source: Michael Everson and Pravin Satpute****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2006-07-03**

This document requests four additional characters to be added to the UCS and contains the proposal summary form. These letters are used in Devanagari orthography for the Kashmiri language, to represent the vowels /i/ and /i:/ (transliterated as *î* and *ï* in the examples below).

In earlier orthographies, these sounds were written with *candra* and the *u*-matras: कँ *kû*, कूं *kü*. A spelling reform approved 2002-08-04 replaced these spellings with क *kû* and कु *kü*. Unfortunately, it is only recently that users have become aware that the characters need to be added to the UCS.

Users have attempted to to employ U+0A41 ँ GURMUKHI VOWEL SIGN U and U+0A42 ँ GURMUKHI VOWEL SIGN UU as a workaround. The Gurmukhi glyphs are unsuitable however, being only an approximation of the correct glyphs. Moreover, the characters correspond to U+0941 ँ DEVANAGARI VOWEL SIGN U and U+0942 ँ DEVANAGARI VOWEL SIGN UU, and implementations using Pango or Uniscribe fail because Gurmukhi vowel signs do not combine with Devanagari base letters. The Gurmukhi and Devanagari *u*-matras both historically derive from Brahmi *u*-matras; the Kashmiri signs are innovations, based on the Devanagari *candra*.

Both the combining marks and the independent vowels are requested; compare U+0946 ऐ and U+0904 ऐ; U+093E ा and U+0906 आ; U+0949 ँ and U+0911 औ; U+094A ौ and U+0912 औ; U+094B ौ and U+0913 औ; U+094C ौ and U+0914 औ; and U+0945 ँ and U+0972 अँ.

0956	◌ँ	DEVANAGARI VOWEL SIGN UE
0957	◌ूं	DEVANAGARI VOWEL SIGN UUE
0976	अ	DEVANAGARI LETTER UE
0977	अ	DEVANAGARI LETTER UUE

Unicode Character Properties

```
0956;DEVANAGARI VOWEL SIGN UE;Mn;0;NSM;;;;N;;;;;
0957;DEVANAGARI VOWEL SIGN UUE;Mn;0;NSM;;;;N;;;;;
0976;DEVANAGARI LETTER UE;Lo;0;L;;;;N;;;;;
0977;DEVANAGARI LETTER UUE;Lo;0;L;;;;N;;;;;
```

Bibliography

Bhat, Shri Roop Krishen. 2002. *Kashmiri Primer*. Northern Regional Language Centre, Central Institute of Indian Languages.

Kaul, Omkar N. *Dictionary of Kashmiri Proverbs*. Patiala: Indian Institute of Language Study.

Raina, M. K. & Neelam Trakru. 2003 *Basic Reader for Kashmiri Language*. Project ZAAN, joint project of Kashmiri Pandits Association Mumbai and LallaDed Educational and Welfare Trust. Website : www.zaan.net, www.koshur.org

Figures

वैलिव कॉशुर हे छव						
सबख - 3						
अु अू						
अछुर	मात्रा	लफ़ज़				
अु	ः	बु	तु	मु	नु	तुरु
		सुह	रुश	असुल	कसुब	बतु
		गरु	कलु	अथु	नतु	गासु
		तनु	गतु	नखु	कथु	नँखरु
		अँदुर	अँतुर	अनु	करु	परु
		छलु	दरु	थवु	फाकु	
अू	ः	अुठ्मि	तुरु	खुमु	रुफु	
		सुती	दुपमाला			
		आँठ सुह ।	असुल रुफ ।	अँदुर दँज ।		
		कलु छल ।	अथु हाव ।	गरु पख ।		
		बु करु कॉम ।	बु परु अखबार ।	बु छलु दँज ।		
		बतु रन ।	हाख तु अल रन ।	कसुब हाव ।		
		फाकु दर ।	कथु मु कर ।	नँखरु मु कर ।		

Figure 1. Text from Bhat 2002 discussing UE and UUE punctuation.

Northern Regional Language Centre, Central Institute of Indian Languages, (Ministry of HRD, Deptt. of Secondary and Higher Education, Govt. of India), Patiala, took lead in this direction and conducted various seminars and workshops for about 2 years, in which renowned linguists and software experts took part. The Centre, under the leadership of Shri Roop Krishen Bhat, organised another workshop from 20th July 2002 to 4th August 2002 at Jammu, to finalise the script and produce necessary Primer and Reader for the language. The experts who were connected with introducing modified symbols and diacritical marks and finalisation of the script, included Dr. S.N.Bhat Haleem, Dr. S.S.Toshakhani, Dr. O.N.Kaul, Prof H.K.Kaul, Prof. C.L.Sapru, Prof. R.L.Shant, Dr. S.N.Raina, Dr. R.K.Bhat and Dr. R.N.Bhat. It may be mentioned that Shri M.K.Kaw, the then Secretary, Education, Govtt. of India, also showed keen interest in devising this material in the interest of the Kashmiri language.

Following changes have been made by the above Expert Committee in the prevalent Nagari-Kashmiri Alphabet:

1. Avagraha	◌̣	is replaced by	◌̣
2. Combination of	◌̣ ◌̣	is replaced by	◌̣
3. Combination of	◌̣ ◌̣	is replaced by	◌̣
4. Combination of	◌̣ ◌̣	is replaced by	◌̣
5. Combination of	◌̣ ◌̣	is replaced by	◌̣

As a consequence, the Standardised Alphabet for the Devanagari-Kashmiri will be:

Vowels:

अ, अँ, अॄ, आ, आँ, इ, ई, उ, ऊ, ऋ,
ए, ऐ, ओ, औ, अं

Figure 2. Text from Raina and Trakru 2003 describing the spelling reform.

<p>रिवान बालु यारस मोटुय कुक्कलि गू गू - गुगू गू हु हू हू - हू!!</p> <p>(म्वछि चीरिथ) मुम कोस्मन कॉर, थँवम मा गुलन तॉर हुरान खून वन खॉर मे द्राव मोसुमन जुव, ल्वकुट जुव — हु हू हू ल्वकुट जुव — हु हू हू हु हू हू — हू !</p> <p>(सीनु द्वगुनॉविथ) हरुद छुस बु नर्कल, कडान बेकुलन कल मे निश गव कुनुय छल गरा द्रह, गरा ब्यूह — कुनुय ब्यूह — हु हू हू कुनुय ब्यूह — हु हू हू</p>	<p>(युथुय हरद सुंद नचुन तेजान छु, सु छु पनुन्यव जेठचव तु पेत्य दार नमव सुत्य यँम्बुरजलि हुंद ऑव्युल जुव कडनुच कूशिश करान, मगर यँम्बुरजलि हुंजि शह खसु वसि सुत्य युस ठबोम्बुर, बोम्बुरठ आलव छु गछान, सु छु हरद सुंघन इरादन छोलु त्रावान तु तसुंघ ज़ीठ्य तु हॉबतनाक नम छिनु यँम्बुरजलि हुंदिस बदनस नजदीख बँहरान यिनस।) (यि हॉबतनाक म्यवुन तु नचुन येलि कमालस वातान छु, कोहव पँत्य किन्य छे जबरदस वुजमलु तु गगरायि गछान। बुतरात छे त्रन्नान। कुत्य कँट्य छि अलान। पलन छि ज़न गुमु फटान तु च्ववापॉर्य छि ज़न पोहु गट कृहुन ज़ाल वाहरावान। वावु ग्रायन सुत्य सुत्य छि छायि गित्य गछान तु तुफान छु छालु मारान लारान लारान बालु प्यतु ब्वन वसान। अडु मॉर गॉमुन्न तु दुतारस मंज पेमन्न यँम्बुरजल वुछिथ छु तुफान ठाह ठाह करान असान। अमि असनु सुत्य छे ज़न वावस ताम पखु हँरिथ प्यवान।)</p> <p style="text-align: right;">(ब्रॉह कुन जॉरी)</p>
--	---

Figure 3. Text from the journal *Härvan*, 2007, with examples of the characters in use.

A. Administrative

1. Title

Proposal to add four characters for Kashmiri to the BMP of the UCS

2. Requester's name

Michael Everson and Pravin Satpute.

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

Individual contribution.

4. Submission date

2008-07-03

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

Devanagari.

2. Number of characters in proposal

4.

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, Fontographer.

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

Yes.

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?

Mr. M.K. Raina, (Convener of Project Zaan, a joint project of Kashmiri Pandits' Association and Lalla Ded Educational and Welfare Trust, Mumbai; Coauthor of 'Basic Reader for Kashmiri Language' and the 'Information Digest' series published under the Project

Zaan), **Mr. Raman Kaul** (*Koshur Devnagri Editor (Manual)*), *Sharda Font for PCs (Manual | Font)*, *Lalded Font for PCs*), **Mr. Rakesh Pandit** (*Project Administrator for Gnome Kashmiri Translation*)

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?

Speakers of Kashmiri.

4a. The context of use for the proposed characters (type of use; common or rare)

Commonly used for modern texts since 2002.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

In Kashmir.

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

Modern use and accordance with the Roadmap.

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. Similarity to Gurmukhi u and uu is superficial.

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 (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

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?