Proposal for adding four characters in Kashmiri to the BMP of the UCS

Introduction:

This document is the proposal for the inclusion of four additional characters to be added into the UCS for the Kashmiri language.

Rationale behind adding four characters:

The additional characters for the Kashmiri language represent vowels sounds 'û' & 'ü' of Devanagari script. In Devanagari, for Ż sound, vowel is written as ॷ and corresponding matra is written as ॹ while, for sound 'ü', vowel is written as ॸ and matra is written as ॺ. Earlier, the two vowel sounds & corresponding matras were not encoded in the Devanagari code chart, since it was presumed to use nearly similar appearing character shapes of Gurumukhi script. But, using the similar appearing character shapes from another script did not seemed as a feasible solution and thereby till date nobody could successfully implement the Kashmiri language using existing Devanagari Unicode chart (Please see Annex A).

Lot of Kashmiri content is available showing the usage of these characters. For example, periodicals like 'Koshur Samachar, Delhi', 'Härvan monthly net-journal of Project Zaan' and 'Khir Bhawani Times, Jammu' (Please see Annex B for images of these periodicals). Kashmiri-Devanagari writing software (Akruti software) developed by Cyberscape Multimedia Ltd., under the guidance of Project Zaan is also available for users. Additionally, the sounds Ż and ü are represented in Arabic script by Î[u+0625] & Î[u+0673] respectively.
The proposed code point and character properties for Kashmiri language in Devanagari script are as follows:

Proposed code point (additional vowels) for Kashmiri in Devanagari code chart:

U+0977 ढ DEVANAGARI LETTER U WITH CIRCUMFLEX
U+0978 ढ DEVANAGARI LETTER U WITH DIAERESIS
U+0979 ़ DEVANAGARI VOWEL SIGN U WITH CIRCUMFLEX
U+097A ़ DEVANAGARI VOWEL SIGN U WITH DIAERESIS

Unicode character properties:

0977;DEVANAGARI LETTER U WITH CIRCUMFLEX;Lo;0;L;;;;;N;;;;;
0978;DEVANAGARI LETTER U WITH DIAERESIS;Lo;0;L;;;;;N;;;;;
0979;DEVANAGARI VOWEL SIGN U WITH CIRCUMFLEX;Mn0;NSM;;;;;N;;;;;
097A;DEVANAGARI VOWEL SIGN U WITH DIAERESIS;Mn0;NSM;;;;;N;;;;;

Bibliography:

1) Kashmiri Primar' by Shri Roop Krishen Bhat, NORTHERN REGIONAL LANGUAGE CENTRE, Central Institute of Indian Languages, Mysore, 11 December, 2002.
   Link

   Website: www.zaan.net, www.koshur.org

3) 'Dictionary of Kashmiri Proverbs' by Omkar N. Kaul, Indian Institute of Language Study.
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. C.N.Kaul, Prof. H.K.Kaul, Prof. C.L.Sapru, Prof. R.L.Shant, Prof. 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, Govt. 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 \( \mathbb{S} \) is replaced by Ardhachandra \( \mathbb{ammable} \)
2. Combination of \( \mathbb{B} \) is replaced by
3. Combination of \( \mathbb{C} \) is replaced by
4. Combination of \( \mathbb{D} \) is replaced by
5. Combination of \( \mathbb{E} \) is replaced by

Figure 1: From Basic Reader for Kashmiri Language

Vowels:

\[ \text{अ, आ, ओ, ठ, आ, ऑ, छ, ट, ट्यु, दू, डी, ड्र, ए, ऐ, ए, ओ, ऑ, ओ, घं} \]

Figure 2: From Basic Reader for Kashmiri Language

Devanagari-Kashmiri standardization happened in July-August 2002 under the leadership of Shri Roop Krishen Bhat. The standardized alphabet for the Devanagari-Kashmiri is given in Figure 2 above. Characters highlighted within the red rectangle are not incorporated in the Unicode 5.0 Devanagari code chart. \( \mathbb{F} \) is added in the Unicode 5.1.0 Beta standard while \( \mathbb{G}, \mathbb{H}, \mathbb{I}, \mathbb{J} \) and \( \mathbb{K} \) are to be incorporated in the Unicode.
Figure 3: From Book Kashmiri Primar written by Shri Roop Krishna Bhat. Shows vowels and corresponding matras.

\[ \text{अ} \quad \Rightarrow \text{written as अ when first letter of a word, as in अहूद उहुद (enemy / chasers)} \]

\[ \Rightarrow \text{written as अ when not first letter of a word, as in गुर गुर (sudden display of energy), बु बु (I, me), खुर खुर (slip)} \]

\[ \text{अु} \quad \Rightarrow \text{written as अु when first letter of a word, as in अुहूब उहुब (eight times)} \]

\[ \Rightarrow \text{written as अु when not first letter of a word, as in कुल कुल (how many?), तुर तुर (cold)} \]

Figure 4: Shows the usage of characters for vowel sound 'ु' & 'ू' from Basic Reader for kashmiri Language.
Annex A

Kashmiri-Devanagari script got standardized in 2002. Due to unavailability of the proposed characters in Devanagari script many users used Gurumukhi character 'u' & 'uu' matra instead of the proposed characters. It created lots of problems since Gurumukhi characters were specifically designed for Gurumukhi script and it was not matching with the proportions of devanagari script characters. Since it is possible to use Gurumukhi characters for Kashmiri-Devanagari in proprietary encoding format. Same is presently not possible with Unicode, since rendering engine comes into picture for processing Unicode string. People from Kashmiri community tried to used these characters in Unicode, but approach was bit different. Usage of Devanagari letter U+0952 \( \ddot{u} \) anudatta for Kashmiri character \( \ddot{u} \) is observed but is not a recommended measure.

Observations and Inferences:

1. The Gurumukhi shape it not same as required in for Kashmiri language. See Fig. 5 and 6.

\[
\begin{array}{c}
\text{U} & \ddot{u} & \text{U}41 \\
\ddot{u} & \text{U} & \ddot{u}
\end{array}
\]

Figure 5 : Comparison of characters from Kashmiri Primar Book and Unicode Gurumukhi Chart.

lohit punjabi font

Figure 6 : Text Composed using Lohit Punjabi Font.
From Figure 5 and Figure 7, we can view the difference between the two character shapes and their usage. Comparison table below reflects the same:

<table>
<thead>
<tr>
<th>Gurumukhi ੁੁ, ੁੂ</th>
<th>Kashmiri ॹ, ॺ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontally very straight</td>
<td>Matras bent in middle, like devanagari Chandrabindu u+0903</td>
</tr>
</tbody>
</table>

2. In Gurumukhi, ੁ represents sound 'u' while ੂ represents sound 'uu'. Kashmiri characters ॹ and ॺ represents sounds 'û' and 'ü'.
3. Using characters of Gurumukhi script in Devanagari script is very problematic for NLP Operations. Identification of Script will be problematic.

4. Rendering engine (Pango, ICU & Uniscribe) doesn't support mixing of script for one language, as they treat each script differently. They are treating these mix characters of different script as different syllable of Bramhi Script and through dotted circle (U+25cc) showing it as invalid combination. Example, कੁ. Rendering engine needs to assume devanagari syllable can have gurumukhi characters which is wrong.

5. Using these two matras will not solve the problem since Kashmiri language require corresponding character for vowel ढ़ औं. Since as per rule of Bramhi script syllable we can't combine matras to the vowels. Therefore, as per this proposal we need four code points for characters ढ़, ॐ, औं, and ा
Annex B
Images from famous periodicals showing uses of characters proposed

Figure 8 & 9: From Koshur Samachar December 2007.
वस्त्र रज्जाक आसिहे वाँसि योहय पंत्राह दुवंजाह हू। मे कुन वन दोल नज़राह तु वोनुन, “वन तु वन्य क्या बनि यथ वन्यनन। कटि अनु बु आभिमत अम्लकेंट्न चुन्नन तु वन्य लागा। रटु छन्ना लुकन नाह गोमुत। कॉर्डिस ति हरगाह फुटमुन मैर हन मैल्य, दपान छि बी आर्स मे फलॉन्य आशुनावन व्यलायतु प्यटु या अरब प्यटु अन्यमूह। हता यारु, वैति छन्ना असल असल गाँरि मेलान! बा खड़ा हज, रायू वॉन्न मंजूर अहमुदस नॅव गँर। पि छनस मामन करॉंची प्यटु सूजमुत। हता यारु क्या वन्य ? गँर छा किनु क्या ? टिक टिक छूस तिथु पॉठच गँड़ान जळ छि कुकिल वूल्य करान। प्यटु वन्य अख सिरु कथ। मे वोनुनय मंजूर अहमुदन्त्न मामन जि करॉंची हुंज़ुन्त्व गच्चन छि अम्लकेंट्न वॉर्ल्य पनुन नाव लेखान तु कुनान। हयो मंजूराणा…” वस्त्रन लोय मंजूरस नाद, मगर सु आवु तु केंह। वस्त्रन वोन मे ल्याति पॉठच, “सु आसी यारन निश गोमुत नॅव गँर हावनि।” मे कोर कल्त सूल्य तसंज़ि प्रथ कथि आंकार। मे आर्स ना पनुन्य गँर शोरुनावन्य! बु ओसुस दिल।
Acknowledgments:

1) Mr. M.K. Raina

   -- Convener of Project Zaan, a joint project of Kashmiri Pandits’ Association and Lalla Ded Educational and Welfare Trust, Mumbai.
   -- Co-author of ‘Basic Reader for Kashmiri Language’ and the 'Information Digest' series published under the Project Zaan.

2) Mr. Raman Kaul

   -- Koshur Devnagri Editor (Manual), Sharda Font for PCs (Manual | Font), Lalded Font for PCs - Designed by Raman K. Kaul

3) Mr. Rakesh Pandit

   -- Project Admin for Gnome Kashmiri Translation