

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹

Please fill all the sections A, B and C below.

(Please read Principles and Procedures Document for guidelines and details before filling this form.)

See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html> for latest Form.

See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for latest Principles and Procedures document.

See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest roadmaps.

A. Administrative

1. Title:	Proposal to add one combining diacritic to the UCS
2. Requester's name:	Andrew Glass, Stefan Baums, Richard Salomon
3. Requester type (Member body/Liaison/Individual contribution):	Individual contribution
4. Submission date:	8 October 2002
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):	No
Proposed name of script: _____	
b. The proposal is for addition of character(s) to an existing block:	Yes
Name of the existing block:	Combining Diacritical Marks
2. Number of characters in proposal:	1
3. Proposed category (see section II, Character Categories):	Combining
4. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000):	3
Is a rationale provided for the choice?	_____
If Yes, reference: _____	
5. 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 ISO/IEC 10646-1: 2000?	Yes
b. Are the character shapes attached in a legible form suitable for review?	Yes
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? _____	
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: Andrew Glass, http://staff.washington.edu/asg/Downloads/GU.zip	

7. 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
8. 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)?	
The character behaves just like existing combining diacriticals	
9. 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/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.	

¹ Form number: N2352-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain _____	No
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.)? If YES, with whom? British Library/University of Washington Early Buddhist Manuscripts Project If YES, available relevant documents: see attachments	Yes
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference: _____	Scholars
4. The context of use for the proposed characters (type of use; common or rare) Reference: _____	Scholarly; Rare
5. Are the proposed characters in current use by the user community? If YES, where? Reference: Scholars worldwide	Yes
6. After giving due considerations to the principles in <i>Principles and Procedures document</i> (a WG 2 standing document) must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference: There is just one and it should be kept with other combining diacriticals	Yes Yes
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? If YES, is a rationale for its inclusion provided? If YES, reference: _____	No No _____
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference: _____	No _____ _____
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference: See below _____	Yes Yes _____
11. 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)? If YES, is a rationale for such use provided? If YES, reference: see attachment Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference: see attachment	Yes Yes Yes Yes
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	No _____
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference: _____	No _____



Proposal

Since the beginning of the twentieth century over a thousand wooden documents written in the Kharoṣṭhī script have been discovered in Central Asia. The discovery of these documents has led to numerous studies and publications, principally: A. M. Boyer, E. J. Rapson, and E. Senart, *Kharoṣṭhī Inscriptions Discovered by Sir Aurel Stein in Chinese Turkestan*, 3 pts, Oxford: Clarendon Press, 1920–9. There has been renewed interest in these documents following the discovery of the British Library Kharoṣṭhī Fragments (see Salomon 1999). As a result it is necessary to be able to encode the transliteration system used by Boyer, Rapson, and Senart in modern publications in electronic form. There is one diacritical mark used in this transliteration system which is not found in the existing Unicode set of combining diacritical marks, so we propose that it be added to that set as follows:

U+0350		COMBINING DOUBLE RING BELOW
--------	---	-----------------------------

It has been suggested that U+032B COMBINING INVERTED DOUBLE ARCH BELOW (◌̣), U+033C COMBINING SEAGULL BELOW (◌̤) or a combination of the two (◌̥) could be used as approximations. However, it is felt that this would not do justice to the diacritic's independent nature, nor in most implementations result in a satisfactory rendering of it.

The shape for this combining character is based on a visual correspondence with the Kharoṣṭhī diacritical mark it transliterates:

Kharoṣṭhī	Transliteration
	<i>bḥa</i>
	<i>u</i>

Unfortunately, the state of our knowledge of the language of these documents does not permit us to identify the phonetic value of this sign beyond Rapson's suggestion that it may represent a modified vowel sound (Boyer, Rapson, and Senart 1920–9: 299). This diacritic, when used for the transliteration of Kharoṣṭhī, only occurs in combination with the vowels *a* and *u*, namely in the akṣaras: *by*, *bḥa*, *la*, *sa*, *śu*, *u* (see Glass 2000: 37, 43–44).

A separate proposal to encode Kharoṣṭhī in Plane 1 of ISO/IEC 10646 includes a combining double ring below. However, this code point (U+10A0D) is intended for use with the Kharoṣṭhī script, written from right to left. It is desirable to have a distinct code point for the transliteration character despite the similarities in appearance since all other code points required for the transliteration of Kharoṣṭhī are included in the BMP. For font development purposes it is desirable to be able to restrict a transliteration font to coverage of just code points in the BMP. In comprehensive fallback Unicode fonts, a single COMBINING DOUBLE RING BELOW glyph may be unsuitable for combination with both Latin-script and Kharoṣṭhī.

Text Samples

DOCUMENTS FOUND AT RUIN N. XXIV. vi

511. (Plate VIII)

N. xxiv. vi. i. Oblong board. (S. Ser. I, p. 257, Pl. XXV.)

- (Obv.)
- (1) viśudha caçku bhavati prabhasvara viśudha gātra sukhumāla su'chati ciraṃ ca āyu labhati
anālpakaṃ karoṃti yo sātṛaganuktam asya 1² na tasya gaṃḍa³ piṭakā bhavaṃti śi li⁴
pa taṃ ta tri⁵ vi cā⁶ rci kaṃ⁷ nā⁸ suci sugaṃdha⁹ labhati sa āśraya karoṃti yo sātṛa-
ganuktam asya 2 viśāla netra bhavati prabhasvara suvarna gātra abhirūpa darśani
- (2) saṃṣṭa ca pādo utarā prasamṣṭita karoṃti yo sātṛaganuktam asya 3 inasmi dāna vara dāna
uktama jaṃdāka kaṃa śukrya nidarśana tathāgatānaṃ bhavatu hitersina¹⁰ jināna pūjā paramārtha
darśana 4 pratyeka budha ca vivegam āśrta ekābhirāma girikaṃtarālaya¹⁰ svakartha yukta śamathe
gune rata teṣaṃ pi pūjā bhavatu svayambhu[na] 4 1
- (3) koçṇiya gotra prathamana śrāvaka [...] teṣaṃ ca yo āsi subhādra paçima etasmi abhyaṃdare
ye atita jinorasa te pi bhavaṃtu pujita 4 2 imasmi dāna gana¹² rāya utama¹⁰ sthaira ca
madhya navaka ca bhicçku aprāpta tiṣu anubhoṃtu utsuka prāpteṣu bhoṃtu satata viśārata 4
3 imasmi yo bhicçku gane¹³ upasthita

Figure 1. Sample text from the first publication to use COMBINING DOUBLE RING BELOW (Boyer, Rapson, and Senart 1920–9: 185).

G *avai*, *avaci*, is the preterite third-person singular of \sqrt{vac} . It is also attested once in the Niya documents in the phrase *yo aṭha varna abhani muni satomasya so sahasrākṣa avaci* (Boyer, Rapson, and Senart 1920–9: no. 511, rev. 3).⁹³ Although such a form does not appear to be attested in other MIA dialects, both *avaci* and *avacī* are found in BHS (see BHSG, § 32.24). Despite their similarity to the Skt. passive aorist *avāci*, which is acknowledged by Edgerton as a possible source (BHSG, § 32.22), it is preferable to follow Edgerton (BHSG, § 32.24) in seeing BHS *avaci/avacī* and G *avai/avaci* as adaptations of the thematic aorist *avaca* (common in both BHS and Pali) to the *iī* aorists (or, perhaps better, blends of the thematic aorist with the *iī* aorists). An alternative but less

Figure 2. Sample text including COMBINING DOUBLE RING BELOW from a recent publication (Allon 2001: 181).

References

- Allon, Mark. 2001. *Three Ekottarikāgama-Type Sūtras: British Library Kharoṣṭhī Fragments 12 and 14*. Gandhāran Buddhist Texts 2. Seattle: University of Washington Press.
- Boyer, A. M., E. J. Rapson, and E. Senart. 1920–9. *Kharoṣṭhī Inscriptions Discovered by Sir Aurel Stein in Chinese Turkestan*. 3 pts. (pt. 3 by Rapson and P. S. Noble). Oxford: Clarendon Press.
- Glass, Andrew. 2000. “A Preliminary Study of Kharoṣṭhī Manuscript Paleography.” Master’s thesis, Department of Asian Languages and Literature, University of Washington.
- Salomon, Richard. 1999. *Ancient Buddhist Scrolls from Gandhāra: The British Library Kharoṣṭhī Fragments*. Seattle: University of Washington Press.

Comments or Discussion

Please send any responses to this proposal to Andrew Glass (email: asg@u.washington.edu). Please also CC to Richard Salomon (email: rsalomon@u.washington.edu) and Stefan Baums (email: baums@u.washington.edu).