



Universal Multiple Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation

Doc Type:	National Proposal
Title:	Proposal on Adding 3 Tibetan Characters and a symbol for ISO/IEC 10646 in BMP
Source:	China
Status:	To be discussed in WG2
Date:	April 23, 2007

Explanation on Adding




3 Tibetan Characters for ISO/IEC 10646 in BMP

1. Introduction




The 195 Tibetan characters encoded in ISO/IEC 10646, are Tibetan letters, Sanskrit letters, punctuations, astronomy and special symbols. They enable to represent thousands Tibetan characters using dynamic combing methods with minimum code points. However, this encoding scheme is not enough to completely process Modern Tibetan. For example, such as description of glyph and application software development, Meanwhile, because of the standards reasons, it seems still quite difficult to properly solve the problems with Tibetan sorting based on traditional Tibetan order.

In this document, we suggest to add 3 characters to BMP. Since there are 61 free codes to use in the code range U+0F00-U+0FFF. With this addition, we believe the modern can be processed completely.


2. Characters that proposed to add

In this document we propose to add 3 consonants. The 3 consonants that proposed to add are:  (TIBETAN ABOVE TRANSFORMED LETTER RAGU) to 0F6B ,  (TIBETAN ABOVE TRANSFORMED LETTER LA) to 0F6C , and  (TIBETAN ABOVE LETTER SA) to 0F6D .

3. Reasons to add these 3 characters

Character , , and  are above letter of Tibetan composite character. There are three reasons to add these 3 characters:

Firstly, these 3 characters are high frequency character in the Modern Tibetan and usually used as a combining element to form Tibetan composite character, such as ཀ, ཁ, ག, ཇ, ཈, ཉ, ཊ, ཋ, ཌ, ཌྷ, ཎ, ཏ, ཐ, ད, དྷ, ན, པ, ཕ, བ, བྷ, མ, ཙ, ཚ, ཛ, ཛྷ, ཝ, ཞ, ཟ, འ, ཡ, ར, ལ, ཤ, ཥ, ས, ཭, ཯, ཰, ཱ, ི, ུ, ཱུ, ྲྀ, ཷ, ླྀ, ེ, ཻ, ོ, ཽ, ཿ, ེ, ཻ, ོ, ཽ, ཿ etc. These are the basis components of Tibetan composite character.

Secondly, if added these 3 characters, the code of ISO/IEC 10646 in BMP could describe the characteristics of Modern Tibetan character. The Modern Tibetan character is combining of prefix, head letter, root letter, subjoined letter, vowel, suffix and post suffix. The existing ISO/IEC 10646 can't rightly describe these component, for example, the fist code of ག and ཁ all are U+0F66, but  of ག is head letter and ཁ of ཁ is root letter. If added these 3 characters to BMP, from U+0F40 to U+0F69 are the root letters, then it can rightly describe the Modern Tibetan characteristics.

Thirdly, the sort order of Tibetan character base on the root letter, so the rightly

find out root letter of every Tibetan glyph, then Tibetan character has sorted is possible based on BMP.

3. Conclusion

After serious discussion and analysis by Tibetan linguists, encoding experts and software developers in China, all are in favor to add the 3 characters.

ISO/IEC JTC 1/SC 2/2WG 2

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646

A. Administrative

1. Title : Adding 3 Tibetan Characters for ISO/IEC 10646 in BMP
2. Requester's name : China
3. Requester Type (Member body/Liaison/individual contribution): Member body
4. Submission date : 2007-3-17
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 : Tibetan
 - b. The proposal is for addition of character(s) to an existing block: Yes
Name of the existing block: Tibetan
2. Number of characters in proposal : 3
3. Proposed category (see section II ,Character Categories): A
4. Proposed Level of implementation (1, 2 or 3)(see clause 14, ISO/IEC 10646/IEC 10646-1: 2000): 2
5. Is a repertoire including character names provided? Yes
 - a. If yes, are the names in accordance with the 'character naming guideline 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? Tibet University
If available now, identify source(s) for the font (including address, e-mail, ftp-site, etc) and indicate the tools used: No.36 ,Jiangsu road ,Lhasa City,Tibet , P.R.China.
Email: ngodrup@utibet.edu.cn
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 use 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, indexing, transliteration etc. (if yes please enclose information) Yes, see appendix A and B.
9. Additional Information:
Submitters are invited to provide any additional information about Properties of the proposal 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: Casting information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc. Combining behaviour, Default Collection 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 Committee for inclusion in the Unicode Standard.

c. Technical-Justification

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? National Body
If YES, available relevant documents: Info will be provide
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use ,or publishing use)is included? yes
Reference: will provide
4. The context of use for the proposed characters (type of use; common or rare) common
Reference: Yes
5. Are the proposed characters in current use by the user community? Yes
If YES, where? Reference: Worldwide
6. After giving due considerations to the principles in *Principles and Procedures document* (a WG 2 standing document) must the proposed characters be entirely in the BMP? Yes
If YES, is a rationale provided?
If YES, reference: See appendix A.
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? Yes
If YES, is a rationale for its inclusion provided?
If YES, reference: See P2. Reasons of adding these 3 characters
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? Yes
If YES, is a rationale for its inclusion provided?
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 (see clauses 4.12 and 4.14 in ISO/IEC 10646-1:2000)? Yes
If YES, is a rationale for such use provided? Yes
If YES, reference: Will provide
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? Yes
If YES,reference: See P2. Reasons of adding these 3 characters
12. Does the proposal contain characters with any special properties such as control function or similar semantics? Yes
If YES, describe in detail (include attachment if necessary): Describe the characteristics of Modern Tibetan character.
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: _____