

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document**Title: Expert Feedback on the proposed Tangut character set in PDAM 6.2****Source: Michael Everson and Andrew West****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2008-09-24**

This document contains some feedback which has been received from a number of Tangut experts who have a particular interest in the study of Tangut text in digital format. The feedback indicates that these experts do not approve of the repertoire which is under ballot in PDAM 6.2, or of its tacit encoding principles.

From: Viacheslav Zaytsev <slidr76@gmail.com>**Date: 2008-09-07****Subject: Re: [Tangut] Review of Proposed Tangut Repertoire****To: Tangut List <tangut@evertyp.com>**

Dear Andrew,

Thank you very much for the review on the proposed Tangut repertoire. After reading it, checking your arguments about disputed characters in the sources (original Tangut sources published in different time in Russia and China, referenced and not referenced additional scholar works on Tangut) and discussing with other tangutologists around the World I should definitely agree that current proposal on Tangut should be revised more before it will be accepted. Here in Russia we revised this proposal before and helped with some parts of it (with mapping of Kychanov's dictionary, for example) but it's a shame we didn't note these problematic parts. So, I hope it's not too late.

I think that situation when two different (not variants by point of view of other scholars we can find in the literature) Tangut characters are joined in one encoded character is absolutely not acceptable. Although the proposal itself and unbelievable work was done for it should be respected very much.

I hope the consensus on disputed parts of the current Tangut proposal will be found. And I'm happy that it and your work on Tangut radicals will be available one day for needs of all the tangutologists.

Additionally I would like to see if the representative source which was used for the current Tangut proposal (i.e. dissertation of Dr. Han) could be published and easy available for the Tangut scholars for review it in the scientific journals and for refer on it. Practically it's very difficult (almost impossible) to get the access to the original of this work right now. But as I

see it's very reputable, based on many sources and important for studying such a difficult question as Tangut orthography. So it should be available for the scholars as soon as possible.

I think that scanned version of this work can not be enough, especially if we speak about it as about representative source for such an important standard as standard on Tangut encoding.

Best wishes,
Viacheslav Zaytsev
Institute of Oriental Manuscripts of the Russian Academy of Sciences
St.Petersburg
Russia

From: Guillaume Jacques <rgyalrongskad@gmail.com>
Date: 2008-09-03
Subject: Re: [Tangut] Review of Proposed Tangut Repertoire
To: tangut@evertype.com

Dear Andrew,

This proposed set of character would indeed be a catastrophe for Tangutologists.

My general comment would be that we want a Unicode standard to encode texts *as they are*, and if we cannot do this properly for the most important of all Tangut texts (IMHO)—Leilin—there is a major problem in the encoding.

These are quick comments, but in any case I strongly suggest that the uniform encoding for Tangut be revised.

Thank you very much for your excellent scholarship and carefulness.

Best wishes,
Guillaume Jacques
Associate Professor (Linguistics)
Centre de Recherches Linguistiques sur l'Asie Orientale (CRLAO)
Université Paris Descartes
<http://xiang.free.fr>

From: Nathan Hill <nathanwhill@gmail.com>
Date: 2008-09-03
Subject: Re: [Tangut] Tangut Radicals Proposal Version 2
To: tangut@evertype.com

Dear Andrew,

I am more of a Tangut enthusiast than a Tangut scholar, but would still venture to say that the proposals you are making are obvious and essential. Merely the compilation of the data you are preparing in the course of your work is a major contribution to Tangutology. We have seen a number of inconveniences or mistakes be introduced into various parts of Unicode over the years. The time and effort to make a polished, maximally accurate, and usable Tangut encoding is more than worth the effort.

Nathan Hill
Lector in Tibetan
School of Oriental & African Studies
University of London

From: Marc Miyake <amritavira@gmail.com>
Date: 2008-09-03
Subject: Re: [Tangut] Review of Proposed Tangut Repertoire
To: tangut@evertype.com

Dear colleagues,

I am in agreement about the need for changes in the proposal. I wish to express my thanks for N3496 which lays out a concrete argument for reform with many specific examples. I would like to add two more points to strengthen the case for expanding the repertoire:

1. Unification is fine when dealing with a script whose variation is completely understood. However, the issue of whether two Tangut characters are the 'same' or 'different' has not yet been fully resolved to everyone's satisfaction. Han Xiaomang's efforts may not be the last word on this matter. It is risky to base an encoding largely on a single scholar's conclusions, though they may ultimately be correct in many cases. At this point, it might be best to be agnostic and be maximally inclusive with characters as well as 'radicals'.

2. Writing *about* the Tangut script is a major branch of Tangut scholarship. Because the study of the Tangut script is still ongoing, it is especially important to include as many variants as possible so they can be discussed in scholarly works. On my site, I have had to make a few original Tangut character GIFs because they are absent from the Mojikyo font when discussing variation. This problem will be exacerbated by the current proposal which has fewer glyphs than Mojikyo.

Thanks again to all who have contributed to this effort. After ten years of handwriting Tangut, three years of using Mojikyo, and thirteen years of struggling with different numbering systems for Tangut characters, I really want a standard that will last the test of time.

Marc Miyake
www.amritas.com

From: Marc Miyake <amritavira@gmail.com>
Date: 2008-09-10
Subject: Re: [Tangut] Fwd: Tangut Information and Unification
To: tangut@evertype.com

Dear Michael,

Thank you for forwarding Richard Cook's letter. I am disappointed, as I favor the "best plain text encoding model", as John Knightley put it. I appreciate the ability to be able to write sinographic variants in Unicode (e.g., 劍 劍 劍 劍 劍 劍 劍) without resorting to variant selection, and I don't see why adding ~400 more Tangut characters is so horrible, particularly when some of them are not variants at all.

I have been looking at the unified Tangut characters over the past few days. Here's one instance that I thought was particularly disturbing and easily understood by nonspecialists:

U+17248 unifies two antonyms (!):

Li Fanwen 1997 735 *dźjij* R37 1.36 'cool'
not in Kychanov 2006
right side is 'earth' (Kychanov 2006 radical B211)

Li Fanwen 1997 1521 *dźjwīj* R37 1.36 'scorching'
Kychanov 2006: 'heat, scorching, sultry'
but defined by Shi et al's 2000 book on Wenhai baoyun as 清凉 'cool'
right side is 'waist/bird' (Kychanov 2006 radical B234)

Strangely, both graphs have 'fire' on the left even though (at least?) one means 'cool'. I would not be surprised if the meanings and/or readings are incorrect: e.g., is *dźjij* a typo for *dźjwīj*, and did Shi et al. accidentally list the meaning of LFW735 instead of LFW1521? In any case, they have different right sides and they *may* have opposite meanings, so I would rather not take the risk of unifying them.

Marc Miyake
www.amritas.com

From: Guillaume Jacques <rgyalrongskad@gmail.com>
Date: 2008-09-16
Subject: Re: [Tangut] The way forward
To: tangut@evertype.com

Dear Andrew,

As we say in French 'Pourquoi faire simple quand on peut faire compliqué'. I really fail to see any merit in using Variant selectors, this will make using Tangut characters even more complicated than it actually is. It is a pity that Unicode is in the hands of computer technicians that have little actual contact with philologists.

I understand the need of a compromise, but I think we have to try before....

Guillaume

From: Arakawa Shintaro <arakawa@aa.tufs.ac.jp>
Date: 2008-09-24
Subject: Issues with Tangut Encoding Proposal
To: Tangut List <tangut@evertype.com>

Dear Dr. West,

I have had the chance to look at your e-mail regarding the matter of Unicode and Tangut characters. At present, I am quite busy and I apologize that I have not been able to participate in the discussion more fully.

I would like to add my perspective in regard to variant characters in Tangut. As has already been raised in the discussion, there are a number of problems with the treatment of variant characters. For we Tangut philologists, the ability to represent genuine Tangut texts electronically is a matter for rejoicing. It is extremely important for research on Tangut texts that even very small differences in the form of Tangut characters are able to be distinguished. If there are two variant characters A and B, and one character (A) is used in some texts whereas the other (B) is used in another set of texts, it is essential for research on Tangut literature to be able to bring forward and distinguish such cases explicitly. If two variant characters were to unfortunately be united, this would give rise to serious obstacles in the comparison and discussion of texts. For linguists and philologists, when a single character has several variants, I believe that it is important that these variants be separately encoded and not distinguished using Variant Selectors. I hold that it would be beneficial to err on the side of encoding too many distinct characters rather than introducing incorrect unification.

I very much appreciate you having consulted me on this matter, and please do not hesitate to contact me with any further inquiries.

Yours sincerely,

Arakawa Shintaro, Doctor of Letters
Associate Professor
Research Institute of Languages and Cultures of Asia and Africa
Tokyo University of Foreign Studies

親愛なるアンドリュー・ウェスト博士

西夏文字とユニコードの問題に関していくつかのメールを拝見しました。私は最近忙しく、議論に加われなかったことをまずお詫びします。西夏語の異体字 (allomorph) に関する私の意見を述べさせていただきます。議論になっているように西夏語の異体字にはいくつかの問題があります。西夏語文献学者にとって、コンピューター上でテキストが整理されることは喜ばしいことです。しかし特に、テキストの研究にとって、文字の細かい点の相違は非常に重要なものです。異体字AとBがあり、ある文献にはAのみ、ある文献にはBのみ、のような情報のために、異体字の区別は絶対に必要です。もしこれらの字体を統一してしまうと、例のような文献の比較研究に重大な問題が生じます。言語学者、文献学者としては、ある文字にいくつかの異体字がある場合、ヴァリエントセレクターではなく、細かい違いも区別したグリフで表し分ける必要があると思います。異体字を統一して文字数を少なくするよりも、フォントの総数が増えても出来るだけ異体字を用意することが正しい方法だと思います。

東京外国語大学アジア・アフリカ言語文化研究所
准教授 博士
荒川 慎太郎