L2/07-301

/IEC JTC 1/SC 2/WG 2 *N3338*

DATE: 2007-09-16

ISO/IEC JTC 1/SC 2/WG 2

Universal Multiple-Octet Coded Character Set (UCS) - ISO/IEC 10646

Secretariat: ANSI

TITLE:	Response to UC Berkeley's proposals on Tangut
SOURCE:	China
ACTION ID:	To be considered by WG2 experts
DISTRIBUTION:	SC2/WG2 members and Liaison organizations
REFERENCES:	WG2N3297 and 3307

After carefully reviewing the proposals on Tangut (WG2N3297 and 3307), Chinese experts in Beijing discussed and thought that the proposals should be perfected.

1. An introduction to the Tangut script

"Tangut" was an ethnic group appeared in the western China at the mediaeval time. By the last half of the tenth century, the Tangut leader Li Jiqian led his people moved to the Gansu Corridor to exploit their territory. Through a hard warfare about half a century, the Tanguts established the kingdom named the "Great State of White and High" in 1038. Because it located at the west of China, it is called "Western Xia" by the historians. The kingdom governed the area including the whole Ningxia and the parts of Inner Mongolia, Shaanxi, Qinghai and Gansu. Its capital was founded at "Xingqingfu" and "Zhongxingfu" afterwards, now Yinchuan city. After 190 years' existence, the kingdom was destroyed by the army of Chinggis-qan in 1227. Then, its people were scattered and joined into other ethnics. After 200 years, the Tangut language and script vanished at last.

According to the records of Song dynasty, the Tangut script was created in 1036, just two years before the establishment of the kingdom. The script was created by one or a few people in a very short time. But it is surprising that such an extraordinary complex writing expanded rapidly to the whole Corridor in virtue of the powerful influence of the Confucianism and Buddhism, and many literatures were produced with its expanding.

More than 90% of Tangut monuments which preserved nowadays are versions of Buddhism works. The others are very rare. Among them there are various translations from Chinese classics, their dictionaries, law codes, literature works and documents. The cultural relics in the 20th century proved that the Tangut script was in usage till the fifth reign year of Hongzhi in Ming dynasty (1502). Its existence lasted for more than 400 years. After Qing dynasty, it became a died script and was not used anymore.

The Tanguts were affected in a long period by the Chinese culture at the beginning of their kingdom, so they created their writing system based on the Chinese characters. It is said that Tangut script is “looked like Chinese, but incognizant”. The reason is that the creator of Tangut script borrowed the basic strokes of the Chinese but changed their forms. The Tangut script looks much more complicated than Chinese. There are some 6000 Tangut characters discovered so far.

2. Glyphs and font of Tangut for ISO/IEC 10646

What is a standard proposal of the Tangut script? We do not want to discuss the problems on computer technology, but on the academic Tangutology . What is the most important is the precision of the form of the Tangut characters. Then based on the precise form of those characters, we can get a set of scientific codes and reasonable compositor, especially used in publishing practice.

We can see from the practice of the design of a Tangut script repertoire in the recent ten years that the technical problems now can be resolved easily. So how to resolve the academic problems becomes the key to a successful design of a repertoire. And the most important standard to value a repertoire is the deltoid precision. The strokes of Tangut script are heavy and complicated. Some characters are very similar to each other. There are only a little difference between them. It requires a long time of systematic study to learn the differences of them. We can’t assume that all academic problems will be resolved just by a ready character list. And in the process of dealing with the characters by computer, the attitude will bring some warps of transcription. The computer technicians without those knowledge can’t detect these errors at all. So the joining by the Tangutologists holds the balance in the process. A repertoire which is built not only by computer technicians but also by Tangutologists will be of high quality. But a repertoire built just by technicians are always criticized.

According to the actual status, it is unpractical to establish a repertoire of 100% preciseness merely based on the computer technique. In the 200 years history of Xixia kingdom, there are not any works of the script criterion. Naturally, no orthographical documents are preserved. So we have not a set of scientific determinate standard to judge which of the characters are right. Most of the judgments remains to be solved by Tangutologists in future.

Let’s assume what will happen if one person writes the same character for ten times? It is sure that the ten copies will have ten appearances, but none of them will be considered as an error. In the users’ mind, there is a standard about the character, and the standard is not a “point” but a “plane”. It is said that if the copies’ difference can be contained in the field, we consider them as correct ones. The field we are discussing has not an exact boundary. The rule to estimate the copies is whether they will be considered as another character. Take the Chinese script for example, when we write down the word “三” （meaning three）, it is indifferent that the first stroke is longer than the second one or not. But when we write down the word “末”（meaning end）, it is very important that which stroke is longer than the other. Because when the second stroke is longer than the first, the word is no longer “末”, but “未”(meaning have not). The similar conditions exist in the Tangut script. To compare every style of the fonts, we need a large scale of study, and now we can just

depend on the experience of the scholars. If there are not philologists, we will be misled by these similarities.

Moreover, we must consider the styles of the fonts carefully. There are kinds of styles of Tangut script, and they all imitated those of the Chinese fonts. Mostly, the Tangut block-printings and the fine manuscripts adopt the regular script, the roughly manuscripts adopt the running hand and others adopt the cursive hand. Now some Tangut script repertories use two fonts, the “Song script” and the regular script. But according to the cultural relics, the “Song script” was never used in the Tangut monuments. So people have to change the strokes of the regular script into imitation of the “Song script”. It makes the fonts a very strange looking. We have the Chinese “Song script” in the history but such a style of the Tangut script never existed in the history. So most of the Tangutologists consider that we can’t ignore the history and the precision of the script to fabricate some styles like “Song script”. The perfect repertoire ought to include only the regular script and every character is not be transcribed by any modern people but must come from the Tangut originals.

At last, let’s discuss the practice of the script. Now there are some repertories are accepted by the Tangutologists, including those of the Tokyo Foreign language University, Ningxia Academy of Social Sciences, Computer Centre of the Ningxia University, and some ones merely for practice, such as those of the Centre of Tangut Studies of the Ningxia University, Institute of Linguistics of Academia Sinica, and those of the Kyoto University and so on. These repertories are used in publishing. And their practicability is certificated by the scholars.

3. Some obvious mistakes in WG2N3297.

3.1. Wrong glyphs

If you look at the most right glyph, the so called Song style glyph absolutely does not exit in the founded Tangut literatures and is in fact fabricated. The inexistent typeface causes wrong glyphs in the proposal. The right dot is placed at wrong location and the 2 strokes at left corner shall be connected as one stroke.

Apparently, the mistakes are introduced by font designers for the reason of lack of knowledge of Tangut script. Thus, we believe that the assistance of Tangutologists were absent when WG2N3297 was prepared. We still are reviewing the code charts of WG2N3297.

3.2. Inexistent typeface of Tangut shall not be used in ISO/IEC 10646

The so called Song style glyph absolutely does not exit in the founded Tangut literatures and is in fact fabricated. It looks strange by Tangutologists and will not be accepted by the Tangut circle.

3.3. Ordering

WG2N3297 introduces an unusual way of ordering which is not familiar to Chinese Tangut circle. Actually, the Tangut circle is familiar with “four-corner code”（四角號碼）. The ordering not practical and makes Tangutologists feel no way to look up a certain character in code charts.

Besides, when duplicated characters (if found) removed and missing characters (if found) added, the order of current code charts will be modified.

3.4. More mistakes are found and will be raised later.

4. Conclusion

Anyway, at first, a perfect repertoire will contain the most accurate fonts. Then based on the precision of them, its codes and arrangements should be in a more scientific and practical order. Furthermore, the repertoire will be checkout through the practice of publishing. As we known, the key to build up a perfect repertoire is not the technique of computer but the academic level of Tangutology. So we don’t think the proposal WG2N3297 is a mature one. Because it just use some Tangut character lists of some recent works but have not any Tangutologist worked with the computer experts from the beginning to the end. And we never see the repertoire have been used to publish any works.

A new Tangut repertoire is now developed by Chinese Tangutoloists and will be presented to scholars in near future.

For the standardization of encoding Tangut in ISO/IEC 10646, we are pleased to see the contribution from UC Berkeley and we welcome the international cooperation of presenting a more mature proposal.