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

Doc Type:	**Working Group Document**
Title:	**Chinese comments on WG2 N5155**
Source:	**China National Body**
Author:	**Wushour Silamu, Eiso Chan**
Status:	**Member Contribution**
Action:	**For consideration**
Date:	**2021-<mark>02-02</mark>**

After checking the comments in WG2 <mark>N5155</mark> carefully, the Chinese experts provide the feedback comments in this document.

We list our basic attitude as below for the convenience of reading and reviewing.

SN	UCS	Feedback
1	U+0626	disagree
2	U+0641	disagree
3	U+06C5	agree
4	U+0677	only agree to add the annotation
	U+06C7	
5	U+0674	disagree
	U+0675	
	U+0676	
	U+0677	
	U+0678	

1. U+0626

The author suggests adding one annotation for the position of *hamza* (ء) in the isolated form and the final form in Kyrgyz, and modifying the glyphs for the corresponding compatibility characters U+FE89 (ئ) and U+FE8A (ـئ).

China's comments are shown as below.

1) The first important problem is that U+0626 (ئ) is not equal to <U+064A,U+0654> (ي‍) clearly. It is not also to treat U+FE89 (ئ), U+FE8A (ـئ), U+FE8B (ئـ) and U+FE8C (ـئـ) as the compatibility characters of <U+064A,U+0654> (ي‍).

2) There are four corresponding compatibility characters of U+0626 (ئ), which are U+FE89 (ئ), U+FE8A (ـئ), U+FE8B (ئـ) and U+FE8C (ـئـ). This was decided by our discussion at the WG2 meeting.

3) In Kyrgyz, the *hamza* (ء) could be positioned at the top in right or middle for typography. Therefore, it is not better to add the paragraph suggested by the author to UCS and Unicode

Core Spec.

2. U+0641

The author suggests changing the represent use of the letter FEH or the consonant [f] from U+0641 (ف) to U+06A7 (ڧ) in Uyghur, Kazakh and Kyrgyz in the draft version, but Eiso Chan, Kushim Jiang and Kwat Ulan raised their objection to UTC, and the author would put off this issue in another document in future.

China's comments are shown as below.

1) It is true that the glyph of U+0641 (ف) in Uyghur, Kazakh and Kyrgyz is similar to that of U+06A7 (ڧ), but please note that U+0641 (ف) has 4 presentation forms for Uyghur, Kazakh and Kyrgyz, which are different to those of U+06A7 (ڧ). The latter are for other language. Thus, keeping using U+0641 (ف) for the letter FEH or the consonant [f] in Uyghur, Kazakh and Kyrgyz is necessary.

2) There was one 7-bit coded character set of Uyghur in China, GB/T 12050-1989. There are two parts in this standard, the first one is the basic set, and the second one is the supplementary set.

We show the mappings for four presentation forms of the letter FEH or the consonant [f] between this standard and UCS as below.

UCS	Char.	Char. Name	Char. Set	Single 7-bit	Double 7-bit
U+FED1	ف	ISOLATED F	basic set	0x4E	0x324E
U+FED2	ڧ	FINAL F	supplementary set	0x2E	0x332E
U+FED3	ڧ	INITIAL F	supplementary set	0x6E	0x336E
U+FED4	ڧ	MEDIUM F	supplementary set	0x4E	0x334E

As we know, the mappings between the character set standard published or released before ISO/IEC 10646-1:1993 and UCS are important, and it is best to keep them stable for the digital life. In current edition of UCS, there are four compatibility presentation forms for U+0641 (ف), which are suitable to map the four characters in Chinese Uyghur character set; and there is no compatibility presentation form for U+06A7 (ڧ).

3) There are two language use annotation under U+06A7 (ڧ) in UCS, but China suggests removing "Uighur" from here.

3. U+06C5

The author suggests modifying the glyph for U+06C5 (ۅ), adding one annotation for two unifiable glyphs, and updating the Arabic Shaping information.

China's comments are shown as below.

1) The looped form and the barred form are both used in China, and the looped form is the main form as the author pointed out. The author's comment on modifying the glyph is right.

2) It is acceptable to add one annotation for the barred form.

3) The file ArabicShaping.txt is only included in UCD not in the UCS data, but we think the author's comment on updating the information for U+06C5 (ۅ) is acceptable.

4. U+0677 & U+06C7

The author suggests modifying the glyphs for U+0677 (ٷ), U+06C7 (ۇ) and the corresponding compatibility presentation forms which are U+FBD7 (ۇ), U+FBD8 (ۇ), U+FBDD (ٷ), and updating the Arabic Shaping information.

China's comments are shown as below.

1) It is acceptable to add the annotation for the Uyghur, Kazakh and Kyrgyz language uses.

2) We need to remind the author that U+0677 (ٷ) is only used in Kazakh, but U+06C7 (ۇ) is used in Uyghur, Kazakh and Kyrgyz.

3) It is not necessary to modify the glyphs for U+0677 (ٷ), U+06C7 (ۇ) and the corresponding compatibility presentation forms. In China, the *damma* (ُ) is written as the comma-like form (�ۛ) in Uyghur, Kazakh and Kyrgyz everywhere, that means the damma-like form and the comma-like form of *dammas* are unifiable under the unification rule of UCS. If the *damma* forms of the glyphs for U+0677 (ٷ) and U+06C7 (ۇ) were changed, it will mean the damma-like form and the comma-like form are treated as two different symbols or characters in UCS. We noticed that U+06C7 (ۇ) is also used for Azerbaijani based on the language use annotation in UCS, so it is better to ask the corresponding encoding working group and experts from Azerbaijan National Body (AZSTAND).

4) It is not necessary to update the Arabic Shaping information for these two characters. The reasons are that U+06C7 (ۇ) is not only used in China, and the author had not provided the usage in Azerbaijan; there is no precedent on *comma* in ArabicShaping.txt; the comma-like form for damma and the damma-like form for damma are unifiable.

5. U+0674 & U+0675, U+0676, U+0677, U+0678

The author suggests treating U+0674 (ٴ) as a combining mark rather than a spacing character, declaring that U+0675 through U+0678 are the discouraged characters, adding the Jawi language use annotation for U+0674.

China's comments are shown as below.

1) We object to change U+0674 (ٴ) to a combining mark. There is no precedent to change a spacing character to a combining mark because the differences between a spacing character and the corresponding combing mark are not only the glyphs, and the script value and several UCD properties are also different. In Arabic shaping, the visual order of the combining marks is different from other Arabic letters. If a spacing character were changed to the corresponding combining mark, the encoding model has been changed in fact. High *hamza* (ٴ) is a special mark only in Kazakh. We treat U+0675 (ٵ), U+0676 (ٶ), U+0677 (ٷ) and U+0678 (ٸ) as single letters, but when one of them is included in one word, the high *hamza* (ٴ) must be moved to the top right of the whole syllable except U+0643 (ك), U+06AF (گ) and U+06D5 (ە) are included in one word. If the first three letters were used as the final letter of one word, the user should add a high *hamza* (ٴ) in front of the whole character string, at the same time, the final form are the same as U+0627 (ا), U+0648 (و) and U+06C7 (ۇ). If U+0678 (ٸ) were used in one word, the user should also add a high *hamza* (ٴ) in front of the whole character string, at the same time, all the forms are the same as U+0649 (ى). This is the reason why we need to encode a spacing high *hamza* (ٴ).

2) We object to declare that U+0675 through U+0678 are the discouraged characters. And there is no need to change any paragraph related to these four letters in UCS and Unicode.

3) We also noticed the orders of the equivalent sequences of U+0675 (اٴ), U+0676 (وٴ), U+0677 (ۇٴ) and U+0678 (ىٴ) are questionable. The orders of two elements of the equivalent sequences should be interchanged, that means decomposition property values in UCD should be modified.
4) Similar to U+0626 (ئ), the equivalent sequence of U+0678 (ىٴ) should be changed to <U+0674,U+0649> (ٴى) from <U+064A,U+0674> (يٴ) if needed.
5) There are no problems to use the current method for so many years in many companies and industries, so that it is best to keep it stable.

Uyghur, Kazakh and Kyrgyz are treated as the multilingual writing system in China, we must keep the basic forms and the representation forms for letters unique. Thank Ms. Evans for providing the comments, but almost all the comments do not hold water. We hope that she can study these issues more carefully. We can discuss these issues face to face in the next meeting if needed, and we also welcome her to come to China for discussion together.

(End of Document)