

ISO/IEC JTC/SC2/WG2

Universal Multiple—Octet Coded Character Set (UCS)

ISO/IEC JTC/SC2/WG2 N2897

2004.12.25

**TITLE: PROPOSING TO SUPPLEMENT WITH THE SPECIAL SCRIPT
AND CHARACTER FOR UIGHUR, KAZAKH AND KIRGHIZ**

SOURCE: CHINA

STATUS: NATIONAL BODY POSITION

ACTION: CONSIDERATION BY WG2

In order to deal with the incompleteness of information exchanges for Uighur, Kazakh, Kirghiz and Arabic languages and occurrence of double ambiguous code, it is quite necessary to supplement Uighur, Kazakh, Kirghiz letters in table1-2.

As we understand, the nominal forms with all the presentation forms are not entirely identical in both of their variant quantity and shape giving rise to double ambiguous code. In order to deal with ambiguity caused by letters in different language texts, which have the same shape in nominal forms , but it have different quantities or shapes in deformed presentation forms, supplement is necessary.

There are some examples about it in 10646 standards. Such as :

“ ﻱ ” (U+0649) in table 15 and “ ﻱ ” (U+06CC) in table 16;

“ ﻑ ” (U+0641) in table 15 and “ ﻑ ” (U+06A7) in table 16;

“ ﻩ ” (U+0647) in table 15 and “ ﻩ ” (U+06BE) in table 16;

“ ﺀ ” (U+0621) in table 15 and “ ﺀ ” (U+0674) in table 16;

“ ﺀ ” (U+0623) in table 15 and “ ﺀ ” (U+0622) in table 16;

“ ﺀ ” (U+0625) in table 15 and “ ﺀ ” (U+0673) in table 16;

The Nominal form shapes of these letters in pair is the same, but quantity of presentation forms or shapes is different, thus all of them have been supplemented with different nominal form codes. At present, there exist numerous nominal

forms in Uighur, Kazakh and Kirghiz information processing and exchange, which cause ambiguity, since their presentation forms different in quantity or shapes.

1) Table 1-1 lists Uighur, Kazakh , Kirghiz and Arabic letters for nominal form in 10646 and nominal form in original shape is completely different in shape and variant quantity, It is certain produce double ambiguous codes in 10646 Standard. Under sequential number 2 in table 1-1, nominal form “ ئ ” will be supplement to Uighur language , because it has four variants: isolated form (“ ئ ” and “ ڨ ”), final form (“ ئ ” and “ ئ ”)different that Nominal form of Kazakh and Kirghiz letter“ ڨ ”(06d5) which has two variants: isolated form “ ڨ ”, final form “ ئ ”. They are different in shape and variant quantity. If Nominal form of Kazakh and Kirghiz letter “ ڨ ”(06d5) replace to Uighur Nominal form“ ئ ” or isolated form “ ڨ ”, As a result, there appears ambiguity in information exchange.

2)Likewise the nominal form of Arabic, Kazakh, Kirghiz language under sequential number 1,2,3,4,5,6 of table 1-1 are the different in shape which nominal form of Uighur language with variant quantity and part of member in shapes, therefore five nominal forms for Uighur language need to be supplement. For example, under sequential number 3 Arabic letter YEH “ ي ” (U+0649) has two variants: (“ ي ” and “ ي ”), but its corresponding character in Kazakh and Kirghiz has four variants: (“ ي ”, “ ي ”, “ ي ” and “ ي ”), in Uighur has eight variants: , So Arabic letter U+0649 cannot represents Uighur, Kazakh and Kirghiz letter mentioned above. Therefore, Uighur, Kazakh and Kirghiz letter “i” should be separately encoded.

We suggest U+076E as the encoding for Kazakh and Kirghiz letter“ ي ”, U+0772 as the encoding for uighur ي . In addition, U+0648 “ و ” is Arabic,

Kazakh and Kirghiz letter, which has two variants (“ف” and “ڤ”), but it cannot represent letter “ئو” in Uighur language which has four variants: (“ف”, “ڤ”, “ئو”, “ئو”). Although the nominal form of this letter of Arabic, Kazakh, Kirghiz and Uighur languages, it has different variant quantity and shapes. So, it is necessary to encode Uighur letter “ئو” separately, and we suggest (U+076F). Some other letters such as U+06c7 “ئو”, U+06c6 “ئو” and U+0627 “ئا” have the same problem mentioned above. Therefore all of these letters should be encoded separately.

2) We emphasized the importance of reason for the nominal form of Uighur letter “ئا” is non-divisible symbol and it cannot be represent by combining symbol “ئا = ئ + ا”, It is not correspond with uighur letters regulations. From this we can see that shape of Uighur letter “ئا” different from Arabic letter “ا”. It has to be separately encoded. As the same, Uighur letters “ئو”, “ئو”, “ئو”, “ئو” and “ئو” is non-divisible symbol and it cannot be represent by combining symbol with “ئ”. You can see , they are different from Arabic letters “و”, “و”, “و” and “و” in their shape and quantity. Therefore all of these letters need to be encoded separately. Otherwise, these letters will cause double ambiguity.

3) Theirs presentation forms of eight nominal forms for Uighur kazakh and kirghiz language have been ranked among 10646 (see table N. 133 FBEA -- FBFD), bat their nominal forms didn't encoded in the 10646 BMP then.

At present, 06 section under 10646 has been occupied and 07 section were supplemented in some Arabic characters lately. Therefore, the supplement codes of nominal form carried in Uighur, Kazakh and Kirghiz languages have been placed in 07 section under 076B-0772.

Table 1-1: Uighur, Kazakh and Kirghiz Letters that produce to Double Ambiguous Codes in 10646

Sequential number	Code	Language	Nominal form in 10646	Nominal form -in original	Pronunciation	variants								
						Isolated form	Final form	Initial form	Medial form	Isolated form	Final form	Initial form	Medial form	
1	0627	Arabic Kazakh Kirghiz	ا	ا	[a]	ا	ا							
		Uighur	ا	ئا	[a]	ا	ا			ئا	ئا			
2	06D5	Kazakh Kirghiz	е	е	[e]	е	е							
		Uighur	е	ئه	[æ]	е	е			ئه	ئه			
3	0649	Arabic	ى	ى		ى	ى							
		Kazakh Kirghiz	ى	ى	[]	ى	ى	د	د					
		Uighur	ى	ئى	[i]	ى	ى	د	د	ئى	ئى	ئى	ئى	ئى
4	0648	Arabic Kazakh Kirghiz	و	و	[w.o]	و	و							
		Uighur	و	ئو	[o]	و	و			ئو	ئو			
5	06C7	Kazakh Kirghiz	ۇ	ۇ	[u]	ۇ	ۇ							
		Uighur	ۇ	ئۇ	[u]	ۇ	ۇ			ئۇ	ئۇ			
6	06C6	Kazakh	ۆ	ۆ	[v]	ۆ	ۆ							
		Uighur	ۆ	ئۆ	[θ]	ۆ	ۆ			ئۆ	ئۆ			

Table 1-2: Uighur, Kazakh and Kirghiz Letters to Be Supplemented

Language	Final form	Medial form	Initial form	Isolated form	Nominal form	No
Uighur	ا ما FE8E FBEB			ا ئا FE8D FBEA	ئا 076B New	1
Kazakh kirghiz	ه FEEA			ه FEE9	ه 076C New	2
Uighur	ه ه FEEA FBED			ه ه FBEC FEE9	ه 076D New	3
Kazakh kirghiz	ى FEF0	ه FBE9	د FBE8	ى FEEF	ى 076E New	4
Uighur	ى ئى FEF0 FBFA	ئى FBD2 ه FBE9	د ئد FBE8 FBFB	ى ئى FEEF FBF9	ئى 0772 New	5
Uighur	و ئو FEEE FBEE			و ئو FEED FBEE	ئو 076F New	6
Uighur	و ئو FBD8 FBF1			و ئو FBD7 FBF0	ئو 0770 New	7
Uighur	و ئو FBDA FBF3			و ئو FBD9 FBF2	ئو 0771 New	8

Positions we suggest for letters which should be supplemented.

0750

Arabic Supplement

077F

	075	076	077
0	ب 0752	ف 0759	ظ 0770
1	ث 0761	ق 0768	ظ 0771
2	ب 0752	ك 0762	ك 0772
3	ت 0753	ك 0763	
4	ب 0754	ك 0764	
5	ب 0755	م 0765	
6	ن 0756	م 0766	
7	خ 0757	ن 0767	
8	ج 0758	ن 0768	
9	ح 0759	ن 0769	
A	د 076A	ل 076A	
B	ر 076B	ع 076B	
C	س 076C	و 076C	
D	ع 076D	ه 076D	
E	ع 076E	ي 076E	
F	غ 076F	ظ 076F	

FB50

Arabic Presentation Forms-A

	FB5	FB6	FB7	FB8	FB9	FBA	FBB	FBC	FBD	FBE	FBF
0	أ FB60	ز FB69	ق FB70	چ FB80	ک FB90	ن FBA0	م FBB0			و FBE0	و FBF0
1	أ FB61	ذ FB61	ق FB71	چ FB81	ک FB91	ن FBA1	م FBB1			و FBE1	و FBF1
2	ب FB62	ت FB62	ح FB72	د FB82	گ FB92	ا FBA2				و FBE2	و FBF2
3	ب FB63	ث FB63	ح FB73	د FB83	گ FB93	ا FBA3			ك FBD3	و FBE3	و FBF3
4	ب FB64	ث FB64	ج FB74	ذ FB84	گ FB94	ة FBA4			ك FBD4	ي FBE4	و FBF4
5	ب FB65	ث FB65	ج FB75	ذ FB85	گ FB95	ة FBA5			ك FBD5	ي FBE5	و FBF5
6	پ FB66	ط FB66	ح FB76	ذ FB86	گ FB96	ه FBA6			ك FBD6	ي FBE6	ي FBF6
7	پ FB67	ط FB67	ح FB77	ذ FB87	گ FB97	ه FBA7			و FBD7	ي FBE7	ي FBF7
8	پ FB68	ط FB68	چ FB78	ذ FB88	گ FB98	ه FBA8			و FBD8	ي FBE8	ي FBF8
9	پ FB69	ط FB69	چ FB79	ذ FB89	گ FB99	ه FBA9			و FBD9	ي FBE9	ي FBF9
A	پ FB6A	ط FB6A	ح FB7A	ذ FB8A	گ FB9A	ه FBAA			و FBD0A	ي FBEA	ي FBFA
B	پ FB6B	ط FB6B	ح FB7B	ذ FB8B	گ FB9B	ه FBAB			و FBD0B	ي FBEB	ي FBFB
C	پ FB6C	ط FB6C	چ FB7C	ذ FB8C	گ FB9C	ه FBA0C			و FBD0C	ي FBE0C	ي FBFC
D	پ FB6D	ط FB6D	چ FB7D	ذ FB8D	گ FB9D	ه FBA0D			و FBD0D	ي FBE0D	ي FBFD
E	ن FB6E	ق FB6E	ح FB7E	ک FB8E	س FB9E	م FBAE			و FBD0E	و FBE0E	ي FBFE
F	ن FB6F	ق FB6F	ح FB7F	ک FB8F	س FB9F	م FBAF			و FBD0F	و FBE0F	ي FBFF

Proposing to supplement with name and code for Uighur eight letters:

ئا	076B	Arabic letter	ALEF	for	Uighur
ە	076C	Arabic letter	AE	for	Kazakh,Kirghiz
ئە	076D	Arabic letter	AE	for	Uighur
ى	076E	Arabic letter	YEH	for	Kazakh,Kirghiz
ئى	0772	Arabic letter	YEH	for	Uighur
ئو	076F	Arabic letter	WAW	for	Uighur
ئۇ	0770	Arabic letter	U	for	Uighur
ئۆ	0771	Arabic letter	OE	for	Uighur

Uighur alphabet for children (nominal forms of Uighur letters):



Eight Uighur letters which should be supplemented and their variants in words:

ئا : ئالم , ئاستا , ئادالەت , ئىمانە , ئامراق
 ئە : ئەمگەك , ئەقىل , ئەسەت , ئەدەب
 ئى : ئېيىق , سېسىق , دېتال , ئېرلان , ئېرا
 ئى : ئىشك , ئىلمى , رۇقىي , رىۋايەت , كۆزى
 ئو : ئوغاق , روشەن , ئوغلاق , ئوبخاق , ئوبلان
 ئۇ : ئۇيغۇر , ئۇچۇر , ئۇزۇن , مەسئۇل , رۇسۇل
 ئۆ : ئۆدەك , كۆل , گۆرۈ , ئۆي
 ئۆ : ئۆزۈم , ئۈرۈمچى , كۆز , ئۈششۈك

Followings are the pictures from Uighur alphabet for children (eight letters which should be supplemented):

ئا ا ا ئا

 ئارا	 ئاي
 ئاپئاق	 ئار

ئاپام ئايغا ئوخشايدۇ.

پ پ پ پ پ پ پ پ

 پىغىز	 پېتىز
 مۈشۈكپىيىق	 دېغىز

ئېيىق ئورمانغا ئامراق،
كېيىقتىن تۇرغىن يىراق.

ئە ە ئە ئە

 دەرخ	 ئەينەك
 مەشئەل	 كۆخلەك

— كەينەك دېگەن؟
— كەينەك!
— يۈرەمە ھەرگىز مەينەت.

ئى ئى ئى ئى ئى ئى

 ئىككى	 ئىت
 گىلاس	 پىل

ئىككى خوراز سوقۇشنى،
يۈز - كۆرنى چوقۇشنى،
ئىككى تايچاق دوست بولۇپ،
ئەڭ ئويلىدى چېشىنى.

ئو و ئو ئو

 دوپپا	 ئوكۇل
 رادىئو	 قول

ئون پارماتتا ئون ئونراق،
كوك ئولۇم ، سول ئولۇم،
ھەر ئولۇمدا بەش پارمات.

ئۇ ۇ ئۇ ئۇ

 دۈمباق	 ئۇۋا
 مايمۇن	 ئاچقۇچ

ئۇلۇمۇ زىيەكەي،
بۇلۇمۇ زىيەكەي،
پاتلىقلىق مەنەم،
كەيسەدىن كەكەي.

ئۈ ۈ ئۈ ئۈ

 كۈنلۈك	 ئۈزۈم
 گۈل	 بۈركۈت

ئۈزۈم ئۈزۈم ئۈزۈدۈم،
ئۈزۈمنى ئۈزۈم ئۈزۈدۈم.

ئۆ ۈ ئۆ ۈ

 تۆگە	 تۆردەك
 بۆرە	 ئۆي

قانا شاختى،
قاي - قاي - قاي،
سۇدا ئۆردەك،
قاي - قاي - قاي.