ISO/IEC JTC/SC2/WG2

Universal Multiple—Octet Coded Character Set
(UCS)

ISO/IEC JTC/SC2/WG2 N4067

2011.5.15

TITLE: Proposal to Encode Special Scripts and Characters in UCS for Uighur language
SOURCE: China
STATUS: NATIONAL BODY POSITION
ACTION: CONSIDERATION BY WG2
Supercedes: WG2N3919

In order to deal with the incompleteness of information exchanges for Uighur languages , it is quite necessary to supplement **Eight** special scripts of Uighur letters in ISO/IEC 10646 /Unicode(table1-2, below).

Problame 1: **1**) We heve emphasize that nominal form of Uighur character "ئا" (table1-2, below) is a single and indivisible symbol, that is represented by combining symbol "ئا = ﺉ+ﺍ " is does not consist with the **writing regulations** of Uighur letters. Similarly, Uighur character "ئە = ﺉ + ه , ئى = ﺉ + ى , ئو" "ئو" = ﺉ +و, "ﺉﻮ" = ﺉ + وٰ , "ﺉﻮٰ" = ﺉ + ۇ and ﺉ = ﭜﭞ , ﺉ + ى are also indivisible symbols , do not divise 2 parts seem above,   （ peleas see table1-5, table1-6, table1-7 ） . Exclusion ，These characters unallowed be represented by combining symbol with "ﺉ"（other wise, the ﺉ special scripts of Uighur letter has not encoded in 10646 /Unicode till now.）, it is a completely single and a indivisible symbols .

2) For the moment, the **Eight** special scripts of Uighur letters represented **by** dual-joining with ﺉ U+0626 ARABIC LETTER YEH WITH HAMZA ABOVE is a wrong action. Bicose U+0626 ى character cannot substitute ﺉ HAMZA ABOVE, thus are completely differente in the shape and quantity(ﺉ have 4 presentation form xxxx , ﺉ only have 2 presentation form xxxx). Divise 2 parts seem above **1**) "ئا = ﺉ+ﺍ " atc is complitly different with Divise 2 parts seem **2**) and such as NFS below in the shapes and quantity. such as NFS.   ﺉ   thus it well to be separately encoded

"05F7" = ئا = |+ ئ   U+0626 ARABIC LETTER YEH WITH HAMZA ABOVE and
        U+0627 ARABIC LETTER ALEF

"05F8" = ئە = ﺉ+ ه   U+0626 ARABIC LETTER YEH WITH HAMZA ABOVE and
U+06D5 ARABIC LETTER AE

"05FB" =و+ئ= **ئو** U+0626 ARABIC LETTER YEH WITH HAMZA ABOVE and

U+0648 ARABIC LETTER WAW

"05FC" =ۇ+ئ= **ئۇ** U+0626 ARABIC LETTER YEH WITH HAMZA ABOVE and

U+06C7 ARABIC LETTER U

"05FD" = ۆ+ئ = **ئۆ** U+0626 ARABIC LETTER YEH WITH HAMZA ABOVE and
U+06C6 ARABIC LETTER OE

"05FA" = ۇ+ئ = **ئۇ** U+0626 ARABIC LETTER YEH WITH HAMZA ABOVE and

U+0649 ARABIC LETTER ALEF MAKSURA

So These characters represented by sequence of two existing characters is a completely wrong action,

3) At the same time, the Uighur letter "ئا" is different to Arabic letter "ا" in shapes and quantity. Also Uighur character "ئۇ", "ئۆ", "ئۇ", "ئو", "ئى", "ئە", and **ئى** are different to Arabic letters "ە", "ى", "و", "ۆ", "ۇ", "ۇ", "ى" in shapes and quantity (table1-1, table1-5, below).

4) The presentation forms of eight nominal forms for Uighur, languages are already encoded UCS extended (see FBEA -- FBFD in Table 133), but their nominal forms are not encoded till now.

Problem 2: There existing the **Eight** special scripts of Uighur letters with Kazakh , Kirghiz and Arabic in ISO/IEC 10646 /Unicode(table1-1, below) which cause ambiguity for information processing and exchange, since their presentation forms are different in quantity or shapes.

1) Table 1-1 lists Uighur, Kazakh, Kirghiz and Arabic letters for nominal form in UCS and nominal form in original shape, which are completely different in shape and variant quantity. It certainly produces double ambiguous codes in UCS. Under serial number 2 in table 1-1, nominal form "ئە" will be supplemented to Uighur language, because it has four variants: isolated form ("ئە " and " ە "), final form ("ـە" and "ـئە"). While the nominal form of Kazakh and Kirghiz letter"ە"(06d5) has two variants: isolated form "ە" and final form "ـە ". Very clear, If Nominal form of Kazakh and Kirghiz letter "ە"(06d5) replace with Uighur Nominal form"ئە " or isolated form " ە ", there appears ambiguity in information exchange . Because"ە"(06d5) and "ئە " are different in shape and variant quantity, so the Arabic letter "ە"(06d5) be ignorant of select two variants "ە" "ـە ", or four variants "ـئە" "ـە " " ە " " ئە".

2) Like wise the nominal form of Arabic, Kazakh, Kirghiz language under serial number 1,2,3,4,5,6 of table 1-1 are different in shape and variant quantity comparing to nominal forms of Uighur language which variant quantity and shapes,. For example, under serial number 3, Arabic letter YEH "ى" (U+0649) has two variants: ("ى" and "ى"), but its corresponding character in Kazakh and Kirghiz has four variants: ("ى"," ى ","ـ" and "ﮗ"). In Uighur, it has **Eight** variants.

So Arabic letter U+0649 cannot be used to represent Uighur, Kazakh and Kirghiz letter mentioned above. Therefore, Uighur letter "i" "ئى" should be separately encoded , suggestion code points to U+05FA.

In addition, U+0648 "و" is Arabic, Kazakh and Kirghiz letter, which has two variants ("و" and "و"), but it cannot be used to represent letter "ئو" in Uighur language which has four variants :("و", "و", "ئو" and "ئو"). So, the Uighur letter "ئو" is necessary to encode separately, we suggest to encode it at U+05FB. Some other letters such as U+06c7 "ئو", U+06c6 "ئو" and U+0627 "ئا" atc have the same problem mentioned above.

Therefore **Eight** nominal forms for Uighur language (table1-6, table1-7) need to be supplemented, and that should be encoded separately.

The Suggested code points for **Eight** nominal forms of Uighur language have to be in 05XX 05f7-05ff (may be arrange in other code points) .

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

Serial 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	ه	ه	[e]	ه	ﻪ						
		Uighur	ه	ئه	[æ]	ه	ﻪ			ئه	ئه		
3	0649	Arabic	ى	ى		ى	ى						
		Kazakh Kirghiz	ى	ى	[ ]	ى	ى	ﺑ	ﺒ				
		Uighur	ى	ئى	[i]	ى	ى	ﺑ	ﺒ	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	ئوۇ
6	06C6	Kazakh	ۆ	ۆ	[v]	ۆ	ۆ				
		Uighur	ۆ	ئۆ	[θ]	ۆ	ۆ			ئۆ	ئوۆ

Table 1-2: Special Scripts to Be Supplemented in ISO/IEC 10646 /Unicode for Uighur languages

Language	Final form		Medial form	Initial form		Isolated form		Nominal form	No
Uighur	ا FE8E	ئا FBEB				ا FE8D	ئا FBEA	ئا 05F7 New	1
Uighur	ه FEEA	ئە FBED				ە FBEC	ئە FEE9	ئە 05F8 New	2
Uighur	ي FBE5	ئې FBF7	ئېـ FBD1 ـ FBE7	يـ FBE6	ئېـ FBF8	ي FBE4	ئې FBF6	ئې 05F9 New	3
Uighur	ى FEF0	ئى FBFA	ئىـ FBD2 ـ FBE9	ىـ FBE8	ئىـ FBFB	ى FEEF	ئى FBF9	ئى 05FA New	4
Uighur	و FEEE	ئو FBEF				و FEED	ئو FBEE	ئو 05FB New	5
Uighur	ۇ FBD8	ئۇ FBF1				ۇ FBD7	ئۇ FBF0	ئوۇ 05FC New	6

Uighur	FBDA FBF3			FBD9 FBF2	05FD New	7
Uighur	FBDC FBF5			FBDB FBF4	05FE New	8

Table 1-3 The Suggested code points for **Special Scripts** of Uighur languages have to be in 05XX 05f7-05ff (may be arrange in other code points) .

	059	05A	05B	05C	05D	05E	05F
0		ͦ 05A0	ͦ 05B0	| 05C0	א 05D0	נ 05E0	וו 05F0
1	ͦ 0591	ͦ 05A1	ͦ 05B1	ͦ 05C1	ב 05D1	ס 05E1	וי 05F1
2	ͦ 0592	ͦ 05A2	ͦ 05B2	ͦ 05C2	ג 05D2	ע 05E2	יי 05F2
3	ͦ 0593	ͦ 05A3	ͦ 05B3	ͦ 05C3	ד 05D3	ף 05E3	′ 05F3
4	ͦ 0594	ͦ 05A4	ͦ 05B4	ͦ 05C4	ה 05D4	פ 05E4	″ 05F4
5	ͦ 0595	ͦ 05A5	ͦ 05B5	ͦ 05C5	ו 05D5	ץ 05E5	
6	ͦ 0596	ͦ 05A6	ͦ 05B6	] 05C6	ז 05D6	צ 05E6	
7	ͦ 0597	ͦ 05A7	ͦ 05B7	ͦ 05C7	ח 05D7	ק 05E7	ﵷ 05F7 New
8	ͦ 0598	ͦ 05A8	ͦ 05B8		ט 05D8	ר 05E8	ﵸ 05F8 New
9	ͦ 0599	ͦ 05A9	ͦ 05B9		י 05D9	ש 05E9	ﵹ 05F9 New
A	ͦ 059A	ͦ 05AA	ͦ 05BA		ך 05DA	ת 05EA	ﵺ 05FA New
B	ͦ 059B	ͦ 05AB	ͦ 05BB		כ 05DB		ﵻ 05FB New
C	ͦ 059C	ͦ 05AC	ͦ 05BC		ל 05DC		ﵼ 05FC New
D	ͦ 059D	ͦ 05AD	ͦ 05BD		ם 05DD		ﵽ 05FD New
E	ͦ 059E	ͦ 05AE	ͦ 05BE		מ 05DE		ﵾ 05FE New
F	ͦ 059F	ͦ 05AF	ͦ 05BF		ן 05DF		

Table 1-4 Arabic Presebtation Forms

	FB5	FB6	FB7	FB8	FB9	FBA	FBB	FBC	FBD	FBE	FBF
0	أَ FB50	ۣ FB60	قُ FB70	چ FB80	ک FB90	ﮠ FBA0	ﮰ FBB0			و FBE0	ﯰ FBF0
1	أَ FB51	ۡ FB61	قُ FB71	چ FB81	ک FB91	ﮡ FBA1	ﮱ FBB1			و FBE1	ﯱ FBF1
2	پ FB52	ﭢ FB62	چ FB72	ڈ FB82	گ FB92	ﮢ FBA2				ۇ FBE2	ﯲ FBF2
3	پ FB53	ﭣ FB63	چ FB73	ڈ FB83	گ FB93	ﮣ FBA3			ﯓ FBD3	ۇ FBE3	ﯳ FBF3
4	ﭔ FB54	ﭤ FB64	چ FB74	ڌ FB84	گ FB94	ﮤ FBA4			ﯔ FBD4	ﯤ FBE4	ﯴ FBF4
5	ﭕ FB55	ﭥ FB65	چ FB75	ﭕ FB85	گ FB95	ﮥ FBA5			ﯕ FBD5	ﯥ FBE5	ﯵ FBF5
6	پ FB56	ﭦ FB66	ج FB76	ڎ FB86	گ FB96	ﮦ FBA6			ﯖ FBD6	ﯦ FBE6	ﯶ FBF6
7	پ FB57	ﭧ FB67	چ FB77	ڎ FB87	گ FB97	ﮧ FBA7			ۆ FBD7	ﯧ FBE7	ﯷ FBF7
8	ﭘ FB58	ﭨ FB68	چ FB78	ڊ FB88	گ FB98	ﮨ FBA8			ۆ FBD8	ﯨ FBE8	ﯸ FBF8
9	ﭙ FB59	ﭩ FB69	چ FB79	ڍ FB89	گ FB99	ﮩ FBA9			و FBD9	ﯩ FBE9	ﯹ FBF9
A	ﭚ FB5A	ﭪ FB6A	چ FB7A	ڌ FB8A	گ FB9A	ﮪ FBAA			و FBDA	ﯪ FBEA	ﯺ FBFA
B	ﭛ FB5B	ﭫ FB6B	چ FB7B	ﭻ FB8B	گ FB9B	ﮫ FBAB			و FBDB	ﯫ FBEB	ﯻ FBFB
C	ﭜ FB5C	ﭬ FB6C	چ FB7C	ﭼ FB8C	گ FB9C	ﮬ FBAC			ﯜ FBDC	ﯬ FBEC	ﯼ FBFC
D	ﭝ FB5D	ﭭ FB6D	چ FB7D	ﭽ FB8D	گ FB9D	ﮭ FBAD			ﯝ FBDD	ﯭ FBED	ﯽ FBFD
E	ﭞ FB5E	ﭮ FB6E	چ FB7E	ک FB8E	ﭞ FB9E	ﮮ FBAE			ۇ FBDE	ۇ FBEE	ﯾ FBFE
F	ﭟ FB5F	ﭯ FB6F	چ FB7F	ک FB8F	ﭟ FB9F	ﮯ FBAF			ۇ FBDF	ﯯ FBEF	ﯿ FBFF

Table 1-5 Names and suggested code points for the eight letters to be supplemented

ئا	05F7	Arabic letter	ALEF	for	Uighur
ئە	05F8	Arabic letter	AE	for	Uighur
ئى	05F9	Arabic letter	YEH	for	Uighur
ئې	05FA	Arabic letter	YEH	for	Uighur
ئو	05FB	Arabic letter	WAW	for	Uighur
ئۇ	05FC	Arabic letter	U	for	Uighur
ئۆ	05FD	Arabic letter	OE	for	Uighur
ئۈ	05FE	Arabic letter	OU	for	Uighur

Table 1-5 Uighur alphabet for children (nominal forms of Uighur letters):

Table 1-6 Proposed eight special letters in Uighur words:

ئا : ئالىم , ئاستا، ئادالەت، ئەئانە، ئامراق

ئە : ئەمگەك، ئەقىل، ئەسەت، ئەدەب

ئې : ئېپىق، سېسىق، دېتال، ئېرلان، ئېرا

ئى : ئىشىك، ئىلمى، رۆقئى، رۇايەت، كۆزى

ئو : ئوغاق، روشەن، ئوغلاق، ئويغاق، ئويلان

ئۆ : ئۆيغۇر، ئۆچۆر، ئۆزۆن، مە سۆل، رۆسۆل

ئۇ : ئۇدەك، كۇل، گۆرۇ، ئۇي

ئۈ : ئۈزۈم، ئۈرۈمچى، كۈز، ئۈششۈك

Table 1-7   Below are the pictures from Uighur alphabet for children (eight letters which should be supplemented):

كۈنلۈك

ئۈزۈم

گۈل

بۈركۈت

ئۆزۈم ئۈزۈم ئۈزدۈم،
ئۈزۈمنى ئۆزۈم ئۈزدۈم.

14

13

تۆگە

ئۆردەك

بۆرە

ئۆي

قاغا شاختا:
—قاق ـ قاق ـ قاق!
سۇدا ئۆردەك:
—غاق ـ غاق ـ غاق!