

Confused Characters

Due to the use of Combining MADDA and HAMZA

Arabic Script IDN Working Group
Version 2.0 12 April 2008

Revision History:

Date	Version	Changes	Contributors
07 April 2008	1.0	Created	A. Al-Zoman
08 April 2008		<ul style="list-style-type: none">- Add description and comment sections – thanks to Roozbeh- Remove characters that are confusable with normalized combining characters (i.g., Alef + Hamza Above <math>\langle \rangle</math> Alef With Wavy Hamza Above) because they will be handled by the “Confusingly similar character” Process.	A. Al-Zoman Roozbeh
12 April 2008	2.0	Restructure the document by arranging combining characters into groups	A. Al-Zoman

This document lists basic characters that are confused with combining characters (generated using combining MADDAH and HAMZA (U+0653 – U+0655))

- 0653 ARABIC MADDA ABOVE
- 0654 ARABIC HAMZA ABOVE
- 0655 ARABIC HAMZA BELOW

Other combining characters (U+0656 – U+065e) are not discussed in this document which may introduce yet another confused characters.

Combining Characters that are Recognized by Unicode Normalization

ا	+	◌َ	=	آ	is confusing with	آ
U+0627		U+0653		U+0627 U+0653		U+0622
Description: Alef + Madda Above <> Alef With Madda Above						
Comments: Unicode normalization takes care of this						

ا	+	◌ْ	=	أ	is confusing with	أ
U+0627		U+0654		U+0627 U+0654		U+0623
Description: Alef + Hamza Above <> Alef With Hamza Above						
Comments: Unicode normalization takes care of this.						

و	+	◌ْ	=	ؤ	is confusing with	ؤ
U+0648		U+0654		U+0648 U+0654		U+0624
Description: Waw + Hamza Above <> Waw With Hamza Above						
Comments: Unicode normalization takes care of this.						

ا	+	◌ِ	=	إ	is confusing with	إ
U+0627		U+0655		U+0627 U+0655		U+0625
Description: Alef + Hamza Below <> Alef With Hamza Below						
Comments: Unicode normalization takes care of this.						

ي	+	◌ْ	=	يْ يِ يَي	is confusing with	يْ يِ يَي
U+064a		U+0654		U+064a U+0654		U+0626
Description: Yeh + Hamza Above <> Yeh With Hamza Above						
Comments: Unicode normalization takes care of this.						

Combining Characters that are Ligatures

ه	+	◌ْ	=	هْ	is confusing with	هْ
U+06c1		U+0654		U+06c1 U+0654		U+06c2
Description: Heh Goal + Hamza Above <> Heh Goal With Hamza Above						
Comments: Unicode normalization takes care of this.						

ع	+	◌ْ	=	عْ	is confusing with	عْ
U+06d2		U+0654		U+06d2 U+0654		U+06d3
Description: > Yeh Barree + Hamza Above <> Yeh Barree With Hamza Above						
Comments: Unicode normalization takes care of this.						

ه	+	◌ْ	=	هْ	is confusing with	هْ
U+06d5		U+0654		U+06d5 U+0654		U+06c0
Description: AE + Hamza Above <> Heh With Yeh Above						
Comments: Unicode normalization takes care of this.						

Unicode Confusable Combining Characters

ح	+	◌ْ	=	حْ	is confusing with	حْ
U+062d		U+0654		U+062d U+0654		U+0681
Description: Hah + Hamza Above <> Hah With Hamza Above						
Comments: This is Unicode confusable!						

ر	+	◌ْ	=	رْ	is confusing with	New codepoint (v 5.1) can not be typed
U+0631		U+0654		U+0631 U+0654		U+076c
Description: Reh + Hamza Above <> Reh With Hamza Above.						
Comments: This is Unicode confusable!						

ى	+	◌ْ	=	ئ ئ ئ	is confusing with	ئ ئ ئ
U+0649		U+0654		U+0649 U+0654		U+0626
Description: Alef Maksura + Hamza Above <> Yeh With Hamza Above						
Comments: This is a Unicode confusable!						

ى	+	◌ْ	=	ئ ئ ئ	is confusing with	ئ ئ ئ
U+06cc		U+0654		U+06cc U+0654		U+0626
Description: Farsi Yeh + Hamza Above <> Yeh With Hamza Above						
Comments: This is Unicode confusable!						

ه	+	◌ْ	=	ه ه ه	is confusing with	ه
U+0647		U+0654		U+0647 U+0654		U+06c0
Description: Heh + Hamza Above <> Heh With Yeh Above						
Comments: This is Unicode confusable! Just an expanded version of " Heh <> Ae "						

Typo/Style Match

ى	+	◌ْ	=	ئ	is confusing with	ئ
U+0649		U+0654		U+0649 U+0654		U+06d3
Description: Alef Maksura + Hamza above <> Yeh Barree With Hamza Above						
Comments: This may be a Typo match for some locales, like the Arabic language users. Just an expanded version of "Alef Maksura <> Yeh Barree".						

ى	+	◌ْ	=	ئ	is confusing with	ئ
U+06cc		U+0654		U+06cc U+0654		U+06d3
Description: Farsi Yeh + Hamza Above <> Yeh Barree With Hamza Above						
Comments: This may be a Typo match for some locales, like the Arabic/Persian language users. Just an expanded version of "Farsi Yeh <> Yeh Barree"						

ى	+	◌ْ	=	ئ	is confusing with	ئ
U+06cd		U+0654		U+06cd U+0654		U+0626
Description: Yeh with Tail + Hamza Above <> Yeh With Hamza Above						

Comments:
 This may be a Typo match for some locales, like the Arabic language users.
 Just an expanded version of " Yeh With Tail <> Alef Maksura"

ي	+	◌ْ	=	يْ	is confusing with	ئ
U+06cd		U+0654		U+06cd U+0654		U+06d3

Description: Yeh With Tail + Hamza Above <> Yeh Barree With Hamza Above

Comments:
 This may be a Typo match for some locales, like the Arabic/Persian language users.
 Just an expanded version of " Yeh With Tail <> Yeh Barree "

ه	+	◌ْ	=	هَّ هَهْ هِهْ هِءْ	is confusing with	هَّ هِءْ
U+0647		U+0654		U+0647 U+0654		U+06c2

Description: Heh + Hamza Above <> Heh Goal With Yeh Above

Comments:
 Since Heh Goal With Yeh Above is equivalent to Heh Goal (U+06C1)+Hamza Above, this simplifies to "Heh <> Heh Goal", which is a Style match in Arabic and Persian, but it's OK in Urdu.

ه	+	◌ْ	=	هَّ	is confusing with	هَّ هِءْ
U+06be		U+0654		U+06be U+0654		U+06c0

Description: Heh Doachashmee + Hamza Above <> Heh With Yeh Above

Comments:
 This simplifies to "Heh Doachashmee <> Ae", which is again a transitive case with normal Heh, as Heh Doachashmee and Ae don't look similar at all to South Asian or Central Asian users.

ه	+	◌ْ	=	هَّ	is confusing with	هَّ هِءْ
U+06be		U+0654		U+06be U+0654		U+06c2

Description: Heh Doachashmee + Hamza Above <> Heh Goal With Hamza Above

Comments:
 This simplifies to "Heh Doachashmee <> Heh Goal", which is again a transitive case with normal Heh. They may be considered Style matches for Persian, but again, since none of them appear in Persian, ...