



Universal Multiple Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation
Международная организация по стандартизации

Doc Type: Working Group Document
Title: About the Code Table Format
Source: V.S. Umamaheswaran (recording secretary), Michel Suignard (project editors), and contributing editors
Status: Expert contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Related:

For nearly 15 years, the repertoire of the Unicode Standard and ISO/IEC 10646 has been synchronized, and for many years, the fonts, tools and people used in preparing the code tables have been shared. Largely for historical reasons, the presentation of the code tables has used a different format for both standards. This document proposes that WG2 and UTC consider aligning the presentation of the code tables. See the summary of the proposal at the end of this document for details.

Background

Continuing with the existing practice has several undesirable consequences.

1. Despite the carefully synchronized repertoire, the standards appear different to the user – this is particularly undesirable now that 10646 is freely available online
2. Two different but incompatible means of providing informative annotations to the characters are being used – these are difficult to maintain and keep synchronized
3. The two formats represent additional workload for the small group of people who have jointly produced the code tables

There are two main parts for the code tables: the code tables for the unified CJK ideographs and all the other scripts (general code tables). Each presents different issues.

CJK Code Tables

For the unified CJK ideographs, the code tables in the Unicode Standard currently show a single glyph per character and the standard provides source information either in machine-readable format or via an online database. This situation is considered less than

ideal, and the Unicode Editorial Committee is actively pursuing the acquisition of the requisite technology to provide multiple-column code tables that correspond to those published in ISO/IEC 10646. Once this technology is available, joint production of CJK code tables would be feasible.

General Code Tables

For the general code tables, a single font collection and a single nameslist file have been used, which have been maintained jointly by several of the contributing editors and the Unicode Editorial Committee. The normative information (character code positions and character names) are therefore guaranteed to be synchronized. However, ISO/IEC 10646 provides a number of notes about particular characters in Annex P. These must be manually correlated to characters for which the nameslist contains a *. This process is inconvenient for the user, and labor intensive for the editors.

In contrast, the character code charts published for the Unicode Standard provide this information inline. That makes the information easier for the editors to maintain, and is much more convenient for the user. WG2 delegates are familiar with this extended format; it has been used by this contributing editor in his regular “Summary of Repertoire” documents for several years now. The extended format does not provide the row, plane or group numbers, as they are redundant and can be trivially deduced from the character code position (see clause 6.2). It also divides the code tables by block, not by half-row. As experience with the “Summary of Repertoire” documents has shown, these differences have not posed any problems to delegates using the documents to verify the repertoire.

The extended format provides several editorial elements that make reading the code tables and locating the correct characters much easier:

1. Subheaders inside a block group characters by category, such as consonants, vowels, punctuation marks
2. Informative aliases provide commonly used alternative names for characters, overcoming a key limitation when the same character is used by multiple groups for multiple purposes under different traditional names, but carries only a single, formal character name
3. Comments indicate usage information that allows users to choose the correct character when name and shape alone are insufficient
4. Cross references guide the user to related characters of similar appearance or similar name or function

In addition, the extended format provides a handy reference of the decomposition information needed by normalization. Except for the last item which is fixed, all the additional information, while helpful to the user, is purely editorial in nature. For the Unicode Standard, this information is contained in the nameslist and has been maintained by the editor of that file. The same person is already a contributing editor and the same file is already used in producing the ISO/IEC 10646 code tables; however, the additional material is

simply suppressed in the current format, while some of it is laboriously repeated in Annex P by the project editor.

Conclusion

If WG2 were to switch to the extended format, not only would the presentation of the code tables between the two standards better reflect the underlying synchronization of the repertoire, but it would be much more convenient to the users of ISO/IEC 10646. At the same time, there would no longer be a need for the project editor to manually maintain Annex P. Instead, editorial suggestions for annotations would be submitted to the editor in charge of the nameslist.







Appendix 1: The Extended Format

The following presents some additional information about the extended format.


Latin-1 punctuation and symbols

Based on ISO/IEC 8859-1 (aka Latin-1) from here.

00A0  NO-BREAK SPACE




- commonly abbreviated as NBSP
- 0020  space
- 2007  figure space
- 202F  narrow no-break space
- 2060  word joiner
- FEFF  zero width no-break space
- ≈ <noBreak> 0020 


00AE  REGISTERED SIGN



- = registered trade mark sign (1.0)
- 24C7  circled latin capital letter r

Letters





00DF  LATIN SMALL LETTER SHARP S

- = Eszett
- German
- uppercase is "SS"
- in origin a ligature of 017F  and 0073 
- 03B2  greek small letter beta


00E5  LATIN SMALL LETTER A WITH RING ABOVE

- Danish, Norwegian, Swedish, Walloon
- ≡ 0061  030A 

...



01C9  LATIN SMALL LETTER LJ
→ 0459  cyrillic small letter lje
≈ 006C  006A 

....

2118  SCRIPT CAPITAL P
= Weierstrass elliptic function

- actually this has the form of a lowercase calligraphic p, despite its name

....

FE18  PRESENTATION FORM FOR VERTICAL RIGHT WHITE LENTICULAR BRACKET
 PRESENTATION FORM FOR VERTICAL RIGHT WHITE LENTICULAR BRACKET


- misspelling of "BRACKET" in character name is a known defect
- ≈ <vertical> 3017 

Figure 1. Examples of character names and annotations in the extended format

The examples shown above start with an example of a subheader and a notice; followed by a sampling of character entries selected to demonstrate different kinds of annotations. In the actual code charts, most characters do not carry any annotations. In the extended format, lines starting with a bullet are informative comments, including information on typical usage. Lines starting with → are informative cross references to characters that are different, but might be of interest to the reader, perhaps, because they might be better alternatives for the purpose, or for other reasons. Lines starting with = give alternative

common names (informative aliases); where such names were used in pre-merger Unicode 1.0, they are marked with “(1.0)”. Lines starting with \equiv or \approx give the decomposition used in normalization. Finally, lines starting with \otimes give the formal character name aliases which are used where character names have been misspelled. For the convenience of the reader, the extended format provides inline glyphs for all character codes.

The final pages of this document reproduce the Arabic block in the extended format as an example of what the code tables for an entire block would look like in the extended format.

Appendix 2: Summary of Proposal

This proposal requests WG2 to consider adopting the proposed Extended Code Table Format, including a new format for the list of character names, and to make the necessary associated changes to Annex P and elsewhere in the standard.

In order to speed up the transition, WG2 is encouraged to adopt this new format at the earliest available opportunity of processing amendments, for example FPDAM4.

A parallel request to the Unicode Consortium to provide multi-column CJK tables is in progress.

If WG2 adopts the proposal, the following actions would result

- 1) The convener to formally request permission from the Unicode Consortium for the use of the additional material present in the nameslist file.
- 2) The editor to review that all information present in annex P is already adequately covered by existing annotations in the nameslist file, otherwise to make the necessary changes.
- 3) The editor to remove the existing text from Annex P and replace it with information on how to use the new format. This text could be based on Appendix 1 of this document.
- 4) The principles and procedures document to point out the informative nature of the additional information and to define the procedures for further, joint, management of editorial changes and additions to the informative material.
- 5) The editor to update clauses 17 and 33 as necessary.
- 6) The editor to prepare the code tables for FPDAM4 in the new format.

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F
0		0610:		0630: ذ	-	0650:	0660: ٠	0670:	0680: پ	0690: ڈ	06A0: غ	06B0: گ	06C0: ٲ	06D0: ٲ	06E0:	06F0: ٠
1		0611:	0621: ء	0631: ر	0641: ف	0651:	0661: ١	0671: آ	0681: خ	0691: ر	06A1: ف	06B1: گ	06C1: ٲ	06D1: ٲ	06E1:	06F1: ١
2		0612:	0622: آ	0632: ز	0642: ق	0652:	0662: ٢	0672: أ	0682: خ	0692: ر	06A2: ف	06B2: گ	06C2: ٲ	06D2: ٲ	06E2:	06F2: ٢
3		0613:	0623: أ	0633: س	0643: ك	0653:	0663: ٣	0673: إ	0683: ح	0693: ر	06A3: ف	06B3: گ	06C3: ٲ	06D3: ٲ	06E3:	06F3: ٣
4		0614:	0624: و	0634: ش	0644: ل	0654:	0664: ٤	0674: ء	0684: ج	0694: ر	06A4: ف	06B4: گ	06C4: و	06D4: -	06E4:	06F4: ٤
5		0615:	0625: إ	0635: ص	0645: م	0655:	0665: ٥	0675: أ	0685: خ	0695: ر	06A5: ف	06B5: ل	06C5: و	06D5: ه	06E5: و	06F5: ٥
6			0626: ئ	0636: ض	0646: ن	0656:	0666: ٦	0676: و	0686: چ	0696: ر	06A6: ق	06B6: ل	06C6: و	06D6:	06E6: ٲ	06F6: ٦
7			0627: ا	0637: ط	0647: ه	0657:	0667: ٧	0677: و	0687: چ	0697: ت	06A7: ف	06B7: ل	06C7: و	06D7:	06E7:	06F7: ٧
8			0628: ب	0638: ظ	0648: و	0658:	0668: ٨	0678: ئ	0688: ڈ	0698: ر	06A8: ق	06B8: ل	06C8: و	06D8:	06E8:	06F8: ٨
9			0629: ٲ	0639: ع	0649: ى	0659:	0669: ٩	0679: ط	0689: د	0699: ر	06A9: ك	06B9: ن	06C9: و	06D9:	06E9:	06F9: ٩
A			062A: ت	063A: غ	064A: ى	065A:	066A: %	067A: ن	068A: د	069A: ن	06AA: ك	06BA: ن	06CA: ق	06DA:	06EA:	06FA: ش
B	060B: ف	061B: ؛	062B: ث		064B:	065B:	066B: ر	067B: ب	068B: ب	069B: پ	06AB: گ	06BB: ن	06CB: و	06DB:	06EB:	06FB: ض
C	060C: ،		062C: ج		064C:	065C:	066C: ’	067C: ت	068C: ڈ	069C: ش	06AC: ك	06BC: ن	06CC: ى	06DC:	06EC:	06FC: غ
D	060D: ،		062D: ح		064D:	065D:	066D: *	067D: ت	068D: د	069D: ص	06AD: ك	06BD: ن	06CD: ى		06ED:	06FD: ٲ
E	060E: م	061E: ٲ	062E: خ		064E:	065E:	066E: س	067E: پ	068E: ڈ	069E: ض	06AE: ك	06BE: ه	06CE: ى		06EE: ذ	06FE: م
F	060F: ع	061F: ؟	062F: د		064F:		066F: و	067F: ت	068F: ڈ	069F: ظ	06AF: گ	06BF: چ	06CF: ف	06DF:	06EF: ر	06FF: ه

Subtending marks

0600		ARABIC NUMBER SIGN
0601		ARABIC SIGN SANAH
0602		ARABIC FOOTNOTE MARKER
0603		ARABIC SIGN SAFHA

Currency sign

060B		AFGHANI SIGN
------	--	--------------

Punctuation

060C		ARABIC COMMA	<ul style="list-style-type: none"> • also used with Thaana and Syriac in modern text → 002C , comma
060D		ARABIC DATE SEPARATOR	

Poetic marks

060E		ARABIC POETIC VERSE SIGN
060F		ARABIC SIGN MISRA

Honorifics

0610		ARABIC SIGN SALLALLAHOU ALAYHE WASSALLAM	<ul style="list-style-type: none"> • represents sallallahu alayhe wasallam “may God’s peace and blessings be upon him”
0611		ARABIC SIGN ALAYHE ASSALLAM	<ul style="list-style-type: none"> • represents alayhe assalam “upon him be peace”
0612		ARABIC SIGN RAHMATULLAH ALAYHE	<ul style="list-style-type: none"> • represents rahmatullah alayhe “may God have mercy upon him”
0613		ARABIC SIGN RADI ALLAHOU ANHU	<ul style="list-style-type: none"> • represents radi allahu ’anhu “may God be pleased with him”
0614		ARABIC SIGN TAKHALLUS	<ul style="list-style-type: none"> • sign placed over the name or nom-de-plume of a poet, or in some writings used to mark all proper names

Koranic annotation sign

0615		ARABIC SMALL HIGH TAH	<ul style="list-style-type: none"> • marks a recommended pause position in some Korans published in Iran and Pakistan • should not be confused with the small TAH sign used as a diacritic for some letters such as 0679 ت
------	--	-----------------------	--

Punctuation

061B		ARABIC SEMICOLON	<ul style="list-style-type: none"> • also used with Thaana and Syriac in modern text → 003B ; semicolon
061C		<reserved>	
061D		<reserved>	
061E		ARABIC TRIPLE DOT PUNCTUATION MARK	
061F		ARABIC QUESTION MARK	<ul style="list-style-type: none"> • also used with Thaana and Syriac in modern text → 003F ? question mark

Based on ISO 8859-6

0621		ARABIC LETTER HAMZA	→ 02BE ʾ modifier letter right half ring
0622		ARABIC LETTER ALEF WITH MADDA ABOVE	≡ 0627 ٱ 0653 ٱ
0623		ARABIC LETTER ALEF WITH HAMZA ABOVE	≡ 0627 ٱ 0654 ٱ

0624		ARABIC LETTER WAW WITH HAMZA ABOVE	≡ 0648 و 0654 ٱ
0625		ARABIC LETTER ALEF WITH HAMZA BELOW	≡ 0627 ٱ 0655 ٱ
0626		ARABIC LETTER YEH WITH HAMZA ABOVE	≡ 064A ي 0654 ٱ

0627		ARABIC LETTER ALEF	
0628		ARABIC LETTER BEH	
0629		ARABIC LETTER TEH MARBUTA	
062A		ARABIC LETTER TEH	
062B		ARABIC LETTER THEH	
062C		ARABIC LETTER JEEM	
062D		ARABIC LETTER HAH	
062E		ARABIC LETTER KHAH	
062F		ARABIC LETTER DAL	
0630		ARABIC LETTER THAL	
0631		ARABIC LETTER REH	
0632		ARABIC LETTER ZAIN	
0633		ARABIC LETTER SEEN	
0634		ARABIC LETTER SHEEN	
0635		ARABIC LETTER SAD	
0636		ARABIC LETTER DAD	
0637		ARABIC LETTER TAH	
0638		ARABIC LETTER ZAH	
0639		ARABIC LETTER AIN	→ 01B9 ع latin small letter ezh reversed → 02BF ʿ modifier letter left half ring

063A

063B

063C

063D

063E

063F

0640 - ARABIC TATWEEL

- = kashida
- inserted to stretch characters
- also used with Syriac

0641		ARABIC LETTER FEH	
0642		ARABIC LETTER QAF	
0643		ARABIC LETTER KAF	
0644		ARABIC LETTER LAM	
0645		ARABIC LETTER MEEM	
0646		ARABIC LETTER NOON	
0647		ARABIC LETTER HEH	
0648		ARABIC LETTER WAW	
0649		ARABIC LETTER ALEF MAKSURA	<ul style="list-style-type: none"> • represents YEH-shaped letter with no dots in any positional form
064A		ARABIC LETTER YEH	

Points from ISO 8859-6

064B		ARABIC FATHATAN	
064C		ARABIC DAMMATAN	
064D		ARABIC KASRATAN	
064E		ARABIC FATHA	
064F		ARABIC DAMMA	
0650		ARABIC KASRA	
0651		ARABIC SHADDA	
0652		ARABIC SUKUN	<ul style="list-style-type: none"> • marks absence of a vowel after the base consonant • used in some Korans to mark a long vowel as ignored • can have a variety of shapes, including a circular one and a shape that looks like ‘◌’ → 06E1 ◌ arabic small high dotless head of khah

Combining maddah and hamza

0653	◌̣	ARABIC MADDAH ABOVE
0654	◌̤	ARABIC HAMZA ABOVE
0655	◌̥	ARABIC HAMZA BELOW

Other combining marks

0656	◌̇	ARABIC SUBSCRIPT ALEF
0657	◌̈	ARABIC INVERTED DAMMA
0658	◌̉	ARABIC MARK NOON GHUNNA <ul style="list-style-type: none"> • Kashmiri and Baluchi • indicates nasalization in Urdu
0659	◌̊	ARABIC ZWARAKAY <ul style="list-style-type: none"> • Pashto
065A	◌̋	ARABIC VOWEL SIGN SMALL V ABOVE <ul style="list-style-type: none"> • African languages
065B	◌̌	ARABIC VOWEL SIGN INVERTED SMALL V ABOVE <ul style="list-style-type: none"> • African languages
065C	◌̍	ARABIC VOWEL SIGN DOT BELOW <ul style="list-style-type: none"> • African languages
065D	◌̎	ARABIC REVERSED DAMMA <ul style="list-style-type: none"> • Ormuri, African languages
065E	◌̏	ARABIC FATHA WITH TWO DOTS <ul style="list-style-type: none"> • Kalami

Arabic-Indic digits

These digits are used with Arabic proper; for languages of Iran, Pakistan, and India, see the Eastern Arabic-Indic digits at 06F0..06F9.

0660	٠	ARABIC-INDIC DIGIT ZERO
0661	١	ARABIC-INDIC DIGIT ONE
0662	٢	ARABIC-INDIC DIGIT TWO
0663	٣	ARABIC-INDIC DIGIT THREE
0664	٤	ARABIC-INDIC DIGIT FOUR
0665	٥	ARABIC-INDIC DIGIT FIVE
0666	٦	ARABIC-INDIC DIGIT SIX
0667	٧	ARABIC-INDIC DIGIT SEVEN
0668	٨	ARABIC-INDIC DIGIT EIGHT
0669	٩	ARABIC-INDIC DIGIT NINE

Punctuation

066A	٪	ARABIC PERCENT SIGN → 0025 % percent sign
066B	٫	ARABIC DECIMAL SEPARATOR
066C	’	ARABIC THOUSANDS SEPARATOR → 0027 ’ apostrophe → 2019 ’ right single quotation mark
066D	*	ARABIC FIVE POINTED STAR → 002A * asterisk

Archaic letters

066E	ﺀ	ARABIC LETTER DOTLESS BEH
066F	ﻉ	ARABIC LETTER DOTLESS QAF

Point

0670	◌̣	ARABIC LETTER SUPERSCRIPIT ALEF <ul style="list-style-type: none"> • actually a vowel sign, despite the name
------	----	---

Extended Arabic letters

0671	اَ	ARABIC LETTER ALEF WASLA <ul style="list-style-type: none"> • Koranic Arabic
0672	اِ	ARABIC LETTER ALEF WITH WAVY HAMZA ABOVE <ul style="list-style-type: none"> • Baluchi, Kashmiri

0673	اِ	ARABIC LETTER ALEF WITH WAVY HAMZA BELOW <ul style="list-style-type: none"> • Baluchi, Kashmiri
0674	اِ	ARABIC LETTER HIGH HAMZA <ul style="list-style-type: none"> • Kazakh • forms digraphs
0675	اِ	ARABIC LETTER HIGH HAMZA ALEF <ul style="list-style-type: none"> • Kazakh ≈ 0627 اِ 0674 اِ
0676	اِ	ARABIC LETTER HIGH HAMZA WAW <ul style="list-style-type: none"> • Kazakh ≈ 0648 وِ 0674 اِ
0677	اِ	ARABIC LETTER U WITH HAMZA ABOVE <ul style="list-style-type: none"> • Kazakh ≈ 06C7 اِ 0674 اِ
0678	اِ	ARABIC LETTER HIGH HAMZA YEH <ul style="list-style-type: none"> • Kazakh ≈ 064A يِ 0674 اِ
0679	اِ	ARABIC LETTER TTEH <ul style="list-style-type: none"> • Urdu
067A	اِ	ARABIC LETTER TTEHEH <ul style="list-style-type: none"> • Sindhi
067B	اِ	ARABIC LETTER BEEH <ul style="list-style-type: none"> • Sindhi
067C	اِ	ARABIC LETTER TEH WITH RING <ul style="list-style-type: none"> • Pashto
067D	اِ	ARABIC LETTER TEH WITH THREE DOTS ABOVE DOWNWARDS <ul style="list-style-type: none"> • Sindhi
067E	اِ	ARABIC LETTER PEH <ul style="list-style-type: none"> • Persian, Urdu, ...
067F	اِ	ARABIC LETTER TEHEH <ul style="list-style-type: none"> • Sindhi
0680	اِ	ARABIC LETTER BEHEH <ul style="list-style-type: none"> • Sindhi
0681	اِ	ARABIC LETTER HAH WITH HAMZA ABOVE <ul style="list-style-type: none"> • Pashto letter “dze”
0682	اِ	ARABIC LETTER HAH WITH TWO DOTS VERTICAL ABOVE <ul style="list-style-type: none"> • not used in modern Pashto
0683	اِ	ARABIC LETTER NYEH <ul style="list-style-type: none"> • Sindhi
0684	اِ	ARABIC LETTER DYEYH <ul style="list-style-type: none"> • Sindhi
0685	اِ	ARABIC LETTER HAH WITH THREE DOTS ABOVE <ul style="list-style-type: none"> • Pashto
0686	اِ	ARABIC LETTER TCHEH <ul style="list-style-type: none"> • Persian, Urdu, ...
0687	اِ	ARABIC LETTER TCHEHEH <ul style="list-style-type: none"> • Sindhi
0688	اِ	ARABIC LETTER DDAL <ul style="list-style-type: none"> • Urdu
0689	اِ	ARABIC LETTER DAL WITH RING <ul style="list-style-type: none"> • Pashto
068A	اِ	ARABIC LETTER DAL WITH DOT BELOW <ul style="list-style-type: none"> • Sindhi
068B	اِ	ARABIC LETTER DAL WITH DOT BELOW AND SMALL TAH <ul style="list-style-type: none"> • Lahnda
068C	اِ	ARABIC LETTER DAHAL <ul style="list-style-type: none"> • Sindhi
068D	اِ	ARABIC LETTER DDAHAL <ul style="list-style-type: none"> • Sindhi

068E	ذ	ARABIC LETTER DUL • older shape for DUL, now obsolete in Sindhi • Burushaski	06A9	ك	ARABIC LETTER KEHEH • Persian, Urdu, ...
068F	ذ	ARABIC LETTER DAL WITH THREE DOTS ABOVE DOWNWARDS • Sindhi • current shape used for DUL	06AA	ك	ARABIC LETTER SWASH KAF
0690	ذ	ARABIC LETTER DAL WITH FOUR DOTS ABOVE • old Urdu, not in current use	06AB	ك	ARABIC LETTER KAF WITH RING • Pashto • may appear like an Arabic KAF (0643 ك) with a ring below the base
0691	ر	ARABIC LETTER RREH • Urdu	06AC	ك	ARABIC LETTER KAF WITH DOT ABOVE • old Malay
0692	ر	ARABIC LETTER REH WITH SMALL V • Kurdish	06AD	ن	ARABIC LETTER NG • Uighur, Kazakh, old Malay, ...
0693	ر	ARABIC LETTER REH WITH RING • Pashto	06AE	ك	ARABIC LETTER KAF WITH THREE DOTS BELOW • Berber
0694	ر	ARABIC LETTER REH WITH DOT BELOW • Kurdish	06AF	گ	ARABIC LETTER GAF • Persian, Urdu, ...
0695	ر	ARABIC LETTER REH WITH SMALL V BELOW • Kurdish	06B0	گ	ARABIC LETTER GAF WITH RING • Lahnda
0696	ر	ARABIC LETTER REH WITH DOT BELOW AND DOT ABOVE • Pashto	06B1	گ	ARABIC LETTER NGOEH • Sindhi
0697	ر	ARABIC LETTER REH WITH TWO DOTS ABOVE • Dargwa	06B2	گ	ARABIC LETTER GAF WITH TWO DOTS BELOW • not used in Sindhi
0698	ج	ARABIC LETTER JEH • Persian, Urdu, ...	06B3	گ	ARABIC LETTER GUEH • Sindhi
0699	ر	ARABIC LETTER REH WITH FOUR DOTS ABOVE • Sindhi	06B4	گ	ARABIC LETTER GAF WITH THREE DOTS ABOVE • not used in Sindhi
069A	س	ARABIC LETTER SEEN WITH DOT BELOW AND DOT ABOVE • Pashto	06B5	ل	ARABIC LETTER LAM WITH SMALL V • Kurdish
069B	س	ARABIC LETTER SEEN WITH THREE DOTS BELOW	06B6	ل	ARABIC LETTER LAM WITH DOT ABOVE • Kurdish
069C	س	ARABIC LETTER SEEN WITH THREE DOTS BELOW AND THREE DOTS ABOVE • Moroccan Arabic	06B7	ل	ARABIC LETTER LAM WITH THREE DOTS ABOVE • Kurdish
069D	س	ARABIC LETTER SAD WITH TWO DOTS BELOW • Turkic	06B8	ل	ARABIC LETTER LAM WITH THREE DOTS BELOW
069E	س	ARABIC LETTER SAD WITH THREE DOTS ABOVE • Berber, Burushaski	06B9	ن	ARABIC LETTER NOON WITH DOT BELOW
069F	ط	ARABIC LETTER TAH WITH THREE DOTS ABOVE • old Hausa	06BA	ن	ARABIC LETTER NOON GHUNNA • Urdu
06A0	ع	ARABIC LETTER AIN WITH THREE DOTS ABOVE • old Malay	06BB	ن	ARABIC LETTER RNOON • Sindhi
06A1	ف	ARABIC LETTER DOTLESS FEH • Adighe	06BC	ن	ARABIC LETTER NOON WITH RING • Pashto
06A2	ف	ARABIC LETTER FEH WITH DOT MOVED BELOW • Maghrib Arabic	06BD	ن	ARABIC LETTER NOON WITH THREE DOTS ABOVE • old Malay
06A3	ف	ARABIC LETTER FEH WITH DOT BELOW • Ingush	06BE	ح	ARABIC LETTER HEH DOACHASHMEE • Urdu • forms aspirate digraphs
06A4	ف	ARABIC LETTER VEH • Middle Eastern Arabic for foreign words • Kurdish	06BF	ح	ARABIC LETTER TCHEH WITH DOT ABOVE
06A5	ف	ARABIC LETTER FEH WITH THREE DOTS BELOW • North African Arabic for foreign words	06C0	ه	ARABIC LETTER HEH WITH YEH ABOVE = arabic letter hamzah on ha (1.0) = izafet • Urdu • actually a ligature, not an independent letter ≡ 06D5 • 0654 ه
06A6	ق	ARABIC LETTER PEHEH • Sindhi	06C1	ح	ARABIC LETTER HEH GOAL • Urdu
06A7	ق	ARABIC LETTER QAF WITH DOT ABOVE • Maghrib Arabic	06C2	ح	ARABIC LETTER HEH GOAL WITH HAMZA ABOVE • Urdu • actually a ligature, not an independent letter ≡ 06C1 ~ 0654 ه
06A8	ق	ARABIC LETTER QAF WITH THREE DOTS ABOVE • Tunisian Arabic	06C3	ح	ARABIC LETTER TEH MARBUTA GOAL • Urdu

06C4	و	ARABIC LETTER WAW WITH RING
		• Kashmiri
06C5	و	ARABIC LETTER KIRGHIZ OE
		• Kirghiz
06C6	و	ARABIC LETTER OE
		• Uighur, Kurdish, Kazakh
06C7	و	ARABIC LETTER U
		• Kirghiz
06C8	و	ARABIC LETTER YU
		• Uighur
06C9	و	ARABIC LETTER KIRGHIZ YU
		• Kazakh, Kirghiz
06CA	و	ARABIC LETTER WAW WITH TWO DOTS ABOVE
		• Kurdish
06CB	و	ARABIC LETTER VE
		• Uighur, Kazakh
06CC	و	ARABIC LETTER FARSI YEH
		• Arabic, Persian, Urdu, ...
		• initial and medial forms of this letter have dots
		→ 0649 و arabic letter alef maksura
		→ 064A و arabic letter yeh
06CD	و	ARABIC LETTER YEH WITH TAIL
		• Pashto, Sindhi
06CE	و	ARABIC LETTER YEH WITH SMALL V
		• Kurdish
06CF	و	ARABIC LETTER WAW WITH DOT ABOVE
06D0	و	ARABIC LETTER E
		• Pashto, Uighur
		• used as the letter bbeh in Sindhi
06D1	و	ARABIC LETTER YEH WITH THREE DOTS BELOW
		• old Malay
06D2	و	ARABIC LETTER YEH BARREE
		• Urdu
06D3	و	ARABIC LETTER YEH BARREE WITH HAMZA ABOVE
		• Urdu
		• actually a ligature, not an independent letter
		≡ 06D2 و 0654 و

Punctuation

06D4	-	ARABIC FULL STOP
		• Urdu

Extended Arabic letter

06D5	و	ARABIC LETTER AE
		• Uighur, Kazakh, Kirghiz

Koranic annotation signs

06D6	و	ARABIC SMALL HIGH LIGATURE SAD WITH LAM WITH ALEF MAKSURA
06D7	و	ARABIC SMALL HIGH LIGATURE QAF WITH LAM WITH ALEF MAKSURA
06D8	و	ARABIC SMALL HIGH MEEM INITIAL FORM
06D9	و	ARABIC SMALL HIGH LAM ALEF
06DA	و	ARABIC SMALL HIGH JEEM
06DB	و	ARABIC SMALL HIGH THREE DOTS
06DC	و	ARABIC SMALL HIGH SEEN
06DD	و	ARABIC END OF AYAH
06DE	و	ARABIC START OF RUB EL HIZB
06DF	و	ARABIC SMALL HIGH ROUNDED ZERO
		• smaller than the typical circular shape used for 0652 و
06E0	و	ARABIC SMALL HIGH UPRIGHT RECTANGULAR ZERO

06E1	و	ARABIC SMALL HIGH DOTLESS HEAD OF KHAH
		= Arabic jazm
		• presentation form of 0652 و, using font technology to select the variant is preferred
		• used in some Korans to mark absence of a vowel
		→ 0652 و arabic sukun
06E2	و	ARABIC SMALL HIGH MEEM ISOLATED FORM
06E3	و	ARABIC SMALL LOW SEEN
06E4	و	ARABIC SMALL HIGH MADDA
06E5	و	ARABIC SMALL WAW
06E6	و	ARABIC SMALL YEH
06E7	و	ARABIC SMALL HIGH YEH
06E8	و	ARABIC SMALL HIGH NOON
06E9	و	ARABIC PLACE OF SAJDAH
		• there is a range of acceptable glyphs for this character
06EA	و	ARABIC EMPTY CENTRE LOW STOP
06EB	و	ARABIC EMPTY CENTRE HIGH STOP
06EC	و	ARABIC ROUNDED HIGH STOP WITH FILLED CENTRE
06ED	و	ARABIC SMALL LOW MEEM

Extended Arabic letters for Parkari

06EE	و	ARABIC LETTER DAL WITH INVERTED V
06EF	و	ARABIC LETTER REH WITH INVERTED V

Eastern Arabic-Indic digits

These digits are used with Arabic-script languages of Iran, Pakistan, and India (Persian, Sindhi, Urdu, etc.). For details of variations in preferred glyphs, see the block description for the Arabic script.

06F0	٠	EXTENDED ARABIC-INDIC DIGIT ZERO
06F1	١	EXTENDED ARABIC-INDIC DIGIT ONE
06F2	٢	EXTENDED ARABIC-INDIC DIGIT TWO
06F3	٣	EXTENDED ARABIC-INDIC DIGIT THREE
06F4	٤	EXTENDED ARABIC-INDIC DIGIT FOUR
		• Persian has a different glyph than Sindhi and Urdu
06F5	٥	EXTENDED ARABIC-INDIC DIGIT FIVE
		• Persian, Sindhi, and Urdu share glyph different from Arabic
06F6	٦	EXTENDED ARABIC-INDIC DIGIT SIX
		• Persian, Sindhi, and Urdu have glyphs different from Arabic
06F7	٧	EXTENDED ARABIC-INDIC DIGIT SEVEN
		• Urdu and Sindhi have glyphs different from Arabic
06F8	٨	EXTENDED ARABIC-INDIC DIGIT EIGHT
06F9	٩	EXTENDED ARABIC-INDIC DIGIT NINE

Extended Arabic letters

06FA	ش	ARABIC LETTER SHEEN WITH DOT BELOW
06FB	ڄ	ARABIC LETTER DAD WITH DOT BELOW
06FC	غ	ARABIC LETTER GHAIN WITH DOT BELOW

Signs for Sindhi

06FD	ۛ	ARABIC SIGN SINDHI AMPERSAND
06FE	۞	ARABIC SIGN SINDHI POSTPOSITION MEN

Extended Arabic letter for Parkari

06FF	و	ARABIC LETTER HEH WITH INVERTED V
------	---	-----------------------------------