

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

Doc Type: Working Group Document**Title: Proposal to encode eight Arabic characters for Persian and Azerbaijani in the UCS****Source: Michael Everson, Roozbeh Pournader, and Elnaz Sarbar****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2006-10-24**

Request. This document asks for the addition of eight Arabic characters to the UCS. One of these letters is frequently used in at least two different common Azerbaijani orthographies and in studies of the Turkic languages. The other seven were used in earlier Persian orthographies, either for archaic Persian (tenth and eleventh centuries CE) or for classical Persian (eleventh to sixteenth centuries CE). Some of them were also used in orthographies for Persian dialects, mostly surviving in a poetic tradition called *Fahlaviyāt*. These characters are used at present by linguists to discuss orthographies, etymology, dialects, manuscripts, and so on.

If this proposal is adopted, the following characters would exist:

اَ	0616	ARABIC SMALL HIGH ALEF-LAM-YEH * early Persian
سَ	063B	ARABIC LETTER SEEN WITH INVERTED V * early Persian
كَ	063C	ARABIC LETTER KAF WITH TWO DOTS ABOVE * early Persian
تَ	063D	ARABIC LETTER KEHEH WITH TWO DOTS ABOVE * early Persian
پَ	063E	ARABIC LETTER KEHEH WITH THREE DOTS BELOW * early Persian
یَ	063F	ARABIC LETTER FARSI YEH WITH INVERTED V * Azerbaijani
تِ	077E	ARABIC LETTER FARSI YEH WITH TWO DOTS ABOVE * early Persian
ثِ	077F	ARABIC LETTER FARSI YEH WITH THREE DOTS ABOVE * early Persian

with the following properties:

```

0616;ARABIC SMALL HIGH ALEF-LAM-YEH;Mn;230;NSM;;;;N;;;;;
063B;ARABIC LETTER SEEN WITH INVERTED V;Lo;0;AL;;;;N;;;;;
063C;ARABIC LETTER KAF WITH TWO DOTS ABOVE;Lo;0;AL;;;;N;;;;;
063D;ARABIC LETTER KEHEH WITH TWO DOTS ABOVE;Lo;0;AL;;;;N;;;;;
063E;ARABIC LETTER KEHEH WITH THREE DOTS BELOW;Lo;0;AL;;;;N;;;;;
063F;ARABIC LETTER FARSI YEH WITH INVERTED V;Lo;0;AL;;;;N;;;;;
077E;ARABIC LETTER FARSI YEH WITH TWO DOTS ABOVE;Lo;0;AL;;;;N;;;;;
077F;ARABIC LETTER FARSI YEH WITH THREE DOTS ABOVE;Lo;0;AL;;;;N;;;;;

```

The shaping classes for these characters, following the syntax of *The Unicode Standard*'s supplementary file ArabicShaping.txt is as follows:

```

063B; SEEN WITH INVERTED V; D; SEEN
063C; KAF WITH 2 DOTS ABOVE; D; KAF
063D; KEHEH WITH 2 DOTS ABOVE; D; GAF
063E; KEHEH WITH 3 DOTS BELOW; D; GAF
063F; FARSI YEH WITH INVERTED V; D; YEH
077E; FARSI YEH WITH 2 DOTS ABOVE; R; YEH WITH TAIL
077F; FARSI YEH WITH 3 DOTS ABOVE; R; YEH WITH TAIL

```

ARABIC SMALL HIGH ALEF-LAM-YEH, ARABIC LETTER FARSI YEH WITH TWO DOTS ABOVE and ARABIC LETTER FARSI YEH WITH THREE DOTS ABOVE. The exact usage of these dotted letters varies according to different manuscripts, authors, and scribes. In general, they are used to differentiate various forms of the very common Persian suffix written with final *yeh*, which is most commonly pronounced [i:]. The modern suffix derives from several suffixes, such as the Middle Persian *-ig* and the Arabic *-iy*; it has a number of grammatical usages. In early Persian or the dialects of the authors or scribes, the difference between the suffixes may have been phonemic (in some dialects of Afghan Persian two different [i:] sounds still exist, though they are not differentiated orthographically). Other writers have used FARSI YEH WITH TWO DOTS ABOVE and FARSI YEH WITH THREE DOTS ABOVE, or the superscript ARABIC SMALL HIGH ALEF-LAM-YEH, to indicate certain grammatical or stress differences. The FARSI YEH WITH TWO DOTS ABOVE and FARSI YEH WITH THREE DOTS ABOVE only appear at the end of words and are attested only in their final and isolated forms. Accordingly, they are classified as right-joining letters—the same joining class of YEH WITH TAIL. (See Figures 1, 2, 3, and 4.)

ARABIC LETTER SEEN WITH INVERTED V is used in the Persian orthography of the twelfth and thirteenth centuries CE and may represent a variety of the phoneme [s] or a similar sound. In modern Persian, words that have been written with the letter are now pronounced with the sound [s] and are written with U+0633 ARABIC LETTER SEEN. (See Figures 5 and 6.)

ARABIC LETTER KAF WITH TWO DOTS ABOVE, ARABIC LETTER KEHEH WITH TWO DOTS ABOVE, and ARABIC LETTER KEHEH WITH THREE DOTS BELOW were used variously to represent the phoneme [g] or the nasalized [ŋ] of Persian dialects). In later orthographies, these were usually replaced either by U+06AF ARABIC LETTER GAF, U+06AB ARABIC LETTER KAF WITH RING, or U+06AD ARABIC LETTER NG. (See Figures 7, 8, 13, 14, and 15.)

ARABIC LETTER FARSI YEH WITH INVERTED V is used in modern Arabic-script orthographies of Azerbaijani. The two most important modern dictionaries of the language, both by Behzad Behzadi, use this letter to distinguish several phonemes that would all be written as *yeh* in older orthographies of Azerbaijani. This FARSI YEH WITH INVERTED V is used to represent the consonant [j], written < y > in the post-1992 Latin orthography of Azerbaijani. In other orthographies of the language or publications discussing the Turkic languages, the same letter is sometimes used in a different way, to represent the phoneme [u] (written < ı > in modern Latin orthographies of Turkish and Azerbaijani). (See Figures 10, 11, and 12.)

Bibliography.

- Behzadi, Behzad. 2003. (1382 AP). *Farhang Azarbaijani-Farsi* (Azerbaijani-Persian dictionary). Tehran, Farhang Moaser. ISBN 964554582X.
- Behzadi, Behzad. 2004 (1383 AP). *Farhang Farsi-Azarbaijani = Farsca-Azərbaycanca sözlük* (Persian-Azerbaijani dictionary). Tehran: Nakhostin. ISBN 9646716008.
- Heyat, Javad. 1988. *Azerbaijan şafāhi xalq adabiyāti* (Azerbaijan oral literature). Tehran: Kavian.
- Heyat, Javad. 2001 (1380 AP). *Seyri dar tārīx-e zabān va lahje-hā-ye torki*. (A tour of the history of Turkic language and dialects). Second edition. Tehran: Peykan. ISBN 9643281663
- Heyat, Javad, et al. 2001. *Türk Dili Yazı Quralları*. Birinci ve İkinci Türk Dili Ortoqrafi Seminarının Aldığı Qərarlar.
- Khalkhali, Nazila. 1996 (1375 AP). *Barresi-e elmi-e šive-ye xatt-e fārsi* (A scientific study of the Persian orthography). Tehran: Qoqnoos. ISBN 9643110532.
- Matini, Jalal. 1968 (1347 AP). “Tahavvol-e rasm-ol-xatt-e fārsi az qarn-e šešom tā qarn-e sizdahom-e hejri” (The evolution of the Persian orthography from the sixth century to thirteenth century AH), in *Magazine of the Literature Department of Mashhad*, Vol. 4, No. 3, pp 135–162.
- Mayel Heravi, Najib. 1990 (1369 AP). *Naqd va tashih-e motun* (Critique and correction of texts). Mashhad: Astan Qods Razavi, the Foundation for Islamic Studies.
- Ravaghi, Ali. 2005 (1383 AP). *Tafsir-e Qor’ān-e pāk* (An interpretation of the holy Qur’an). Tehran: Samt. ISBN 9644599101.
- Sadeghi, Ali Ashraf. 2003. “Fahlaviyāt-e Eyn-ol-Qozāt-e hamedāni va čand fahlavi-e digar” (Ain-al-Qozāt’s *Fahlaviyāt* and some other *Fahlavis*), in *Iranian Journal of Linguistics*. Vol. 18, No. 1: pp 13–24.
- Anonymous. 1991. “Təntənai gözəl bir şe’r gecəsi” (A magnificent, wonderful night of poetry reading), in *Varliq*, Vol. 13, No. 801, pp 108–109.

Acknowledgments. This project was made possible in part by a grant from the High Council of Information Dissemination of Iran (which funded the FarsiWeb Project, Sharif FarsiWeb, Tehran). The authors also wish to thank Amir Hedayaty of Sharif University of Technology for bringing to their attention the unencoded Azerbaijani letter.

Figures

«ی» وحدت در کلمات مختوم به «ی» گاه به صورتِ امروزی مثل گرمیی،
 سردیی و گاه به صورتِ الی ریز در بالای «ی» وحدت ضبط میشده است: «اما
 این زشت نامی الی باشد.»^۱

Figure 1. Sample from Khalhkali 1996 showing ARABIC SMALL HIGH ALEF-LAM-YEH.

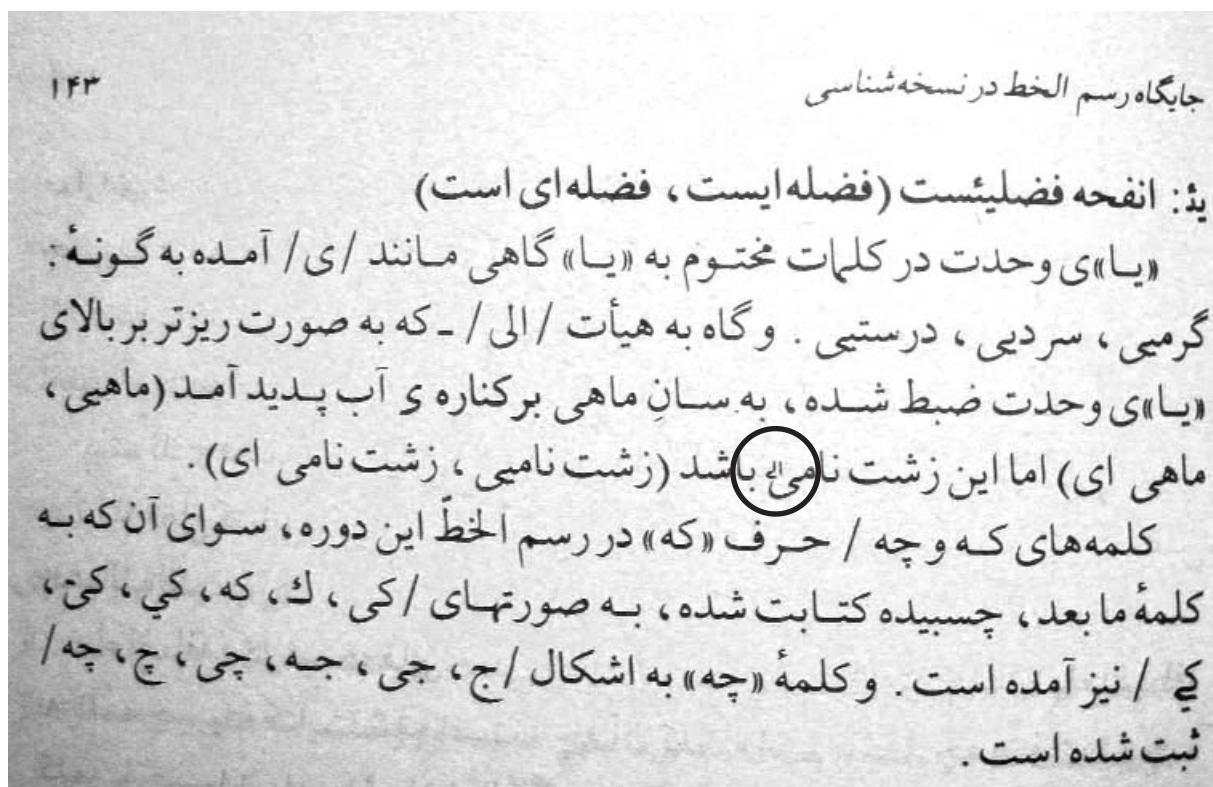


Figure 2. Mayel Heravi 1990 showing ARABIC SMALL HIGH ALEF-LAM-YEH.

«ی» در فارسی انواع گوناگون دارد مثل «ی» وحدت، مجهول، مصدری
و... در نسخه‌های دست نویس گونه‌های نوشتاری زیر دیده میشود:
«ی» چیزی، «ی» کنتی؛ «ی» همی داشتی؛ «ی» ماهی.

Figure 3. Sample from Khalhkali 1996 showing ARABIC LETTER FARSI YEH WITH TWO DOTS ABOVE and ARABIC LETTER FARSI YEH WITH THREE DOTS ABOVE .

تلفظ می کرده‌اند. در سده‌های چهارم و پنجم، در کتاب بیر از عرب و
نشانه‌هایی استفاده می شده که انواع «یا» مشخص شود. متأسفانه به علت از
بین رفتن نسخه‌های کهن از نگاشته‌های فارسی، تاکنون تاریخ گونه‌های زبانی
زبان فارسی و نیز چگونگی تلفظ انواع «یا» بتحقیق برگرفته نشده است، با این
همه /ی/ را به صورتهای ذیل در نسخه‌های بازمانده از دوره مورد بحث
می بینیم:

ی: کردی، چیزی، زیراکي (که) /تی/ کنتی، مردی، موی /یی: همی
داشت، بخوری، هندی /ثی/ ماهی، بیامدی، می جنبانیدی، گفتدی،
اندکی /زیشان، پیمانی، فی.

نشانه اضافه / علامت اضافه در رسم الخط این دوره به صورتهای زیر دیده
می شود:

در کلمات مختوم به «ها» ی غیر ملفوظ، گاه به صورت نیمه /ی/ و به مانند
همزه، و گاه به هیأت /ی/ کامل، و هم گاهی بدون نشانه آمده است مانند
خزانه اوی، زهره عقاب، ریزه او، درجه دوم.

در کلمات مختوم به مصوتهای بلند /ā = آ/ و /u = او/ بیشتر به هیأت /ی/

Figure 4. Sample from Mayel Heravi 1990 showing ARABIC LETTER FARSI YEH WITH TWO DOTS ABOVE and ARABIC LETTER FARSI YEH WITH THREE DOTS ABOVE .

از مختصات دیگر این دوره وجود نقطه و اعراب برای متمایز کردن
حروف همانند است مثل «ر»، «د»، «س»، «گ»، که گاه به صورت «ر»، «د»،
«س»، «ک» نوشته میشود است مانند روا، دبه، تفسیر و گاه به صورت «ر»، «ش»،
ک مثل گوشپند، گفت.

Figure 5. Sample from Khalhkali 1996 showing ARABIC LETTER SEEN WITH INVERTED V.

واژه‌هایی که مختوم به هجای / اوی / بوده است، به هیأت / ا / نویسانیده باشند.

دیگر از خصیصه‌های بارز در رسم الخط قرنهای ششم و هفتم هجری، وجود نقطه و شکل است که کاتبان بر زیر برخی از حروف - به منظور ممتاز کردن آنها از حروف همانند - گذارده‌اند. چنان که حروف / ر، د، س، گ / را در نسخه‌های این دوره گاه به صورت / ر، د، پ، ک، / می‌بینیم. مانند روا، کار، بزرگوار، دنبه، بدو، دهن، بدان، درست، تفسیر، مسلمان، میسون. و گاه سه حرف / ر، س، گ / را به هیأت / ر، س، ک / می‌بینیم. مانند گوسپند، شبک، کتی، کشی (گسی)، کفت (گفت) بازان، بازیدن، زاندن.

کلمه یا حرف / که / نیز در رسم الخط این دوره مانند دوره پیش، گاه به کلمه پس از خود متصل کتابت شده است، مانند کپاره (که پاره‌ای)، کدنبال (که دنبال) کچون (که چون) کتان (که تان). و گاه به صورت جدا، و به صورت

Figure 6. Sample from Mayel Heravi 1990 showing ARABIC LETTER SEEN WITH INVERTED V.

«ک»، در نسخه‌های بازمانده‌ی قرون چهارم و پنجم برای تفارق میان «ک» و «گ» آن را با دو یا سه نقطه در زیر و زیر متمایز کرده‌اند: (ک، گ، ک). حروف «ر»، «د»، «س» را نیز با گذاردن یک نقطه در زیر «د» و «ر» و سه نقطه در زیر «س» مشخص کرده‌اند: پسر، برادر.

Figure 7. Sample from Khalhkali 1996 showing ARABIC LETTER KAF WITH TWO DOTS ABOVE.

فارسی گاهی به صورت /ب و ج/ کتابت شده است، اسباب بسیاری از تصحیفات و تحریفات و سرانجام تصرفات کاتبان را در نسخه‌های خطی فارسی فراهم آورده است.

در نسخه‌های باز مانده از دوره موضوع بحث، در کتابت، میان /گ/ و /ك/ فرق گذارده شده، به نحوی که /گ/ را با دو یا سه نقطه که در زیر آن گذاشته‌اند به این هیأت /كُ، كُ، كُ/ از /ك/ ممتاز کرده‌اند.

ر، د، س/ در برخی از نسخه‌های این دوره، مانند نسخه کراچی از شرح التعریف، مورخ ۴۴۷ هجری، با گذاردن يك نقطه در زیر /ر، د/ و سه نقطه در زیر /س/ مشخص کرده‌اند. مانند بدادر، فرامشت، پیر، سپاعتی

نشانه مدّ/ در رسم الخط این دوره، و نیز به طور پراکنده در کتابت کاتبان سده‌های ششم تا هشتم، نشانه مدّ در کلماتی که پس از حروف اضافه/ به، از، اندر، بر، که/ می آمده، ظاهر نمی شده، و گویا این خصیصه کتابت بر اثر تمایلی بوده است که اهل زبان در نزدیک گردانیدن میان گونه‌های نوشتار و گفتار زبان داشته‌اند. به لحاظ این نکته است که در نسخه‌های باز مانده از

Figure 8. Sample from Mayel Heravi 1990 showing ARABIC LETTER KAF WITH TWO DOTS ABOVE.

اردیبهشت آیین شین بشینده گوزل بیر شعرگجه سینه
هین آذربایجانلی طلبه‌لری، سالی سوزلو طنطنه‌لی بیر شعرگجه سینه
خا برلیق کورموشدولر بو شعرگجه سینه هرندن اؤنجه ایشین منظم و مرتب
لیکی خورما گلن و شویندیریجی ایدی. شعرگجه سینه گنج طلبه و شاعرلردن
باشقا اؤستاد شاعرلردن ده جا غیرمیشدیلا. درگیمیزین یازبیجی و شاعرلر
لریندن مجیدزاده ساوالان و آقای بیت‌الله جعفری ده بو شعرگجه سینه
اشتراک ائتمیشدیلا. طلبه‌لرین شاعر و موسیقی صنعتینده وارد اولانلاریندان
باشقا آذربایجان شین اؤستاد شاعرلریندان با شیق جعفرگوزله‌لینی و بالابا نچی

Figure 9. Sample from Varliq 1991, showing ARABIC LETTER FARSI YEH WITH INVERTED V.

- ۲۶- و : وَاَر، نُو، وُو.
 ۲۷- و : وجوز، بوز، قوزو.
 ۲۸- و : وُز، گولوش، کورُو.
 ۲۹- و : وُت، دُون، بُوغَاَز، کِلُو.
 ۳۰- و : وُيُوش، دُورْد، گُوز.
 ۳۱- و - ي - ي : يَاغ، آيَاق يَامِيَا شِيل، نِيل، آي.
 ۳۲- ي - ي - ي : يِل، دِيل، اِيگِيد، ايرِي.
 ۳۳- ي - ي - ي : دِيرِنَاق، قِيچ، قَاَرشِي، قَاَرِي.

Figure 10. Sample from Behzadi 2003, showing ARABIC LETTER FARSI YEH WITH INVERTED V.

کُيْتِمک { کُيْتِمَاک } باخ: کُيْتِيرِمک.
 کُيْتِمَه { کُيْتِمَا } { ا.ا / ف.ا. } بی-حتی،
 کرختی.
 کُيْتِيرِجِي باخ: کُيْتِيرِجِي.
 کُيْتِيرِشِمک { کُيْتِيرِشِمَاک } باخ: کُيْتِيرِشِمک
 (کُيْتِيرِشِمک).
 کُيْتِيرِکْدِيرِمک { کُيْتِيرِکْدِيرِمَاک } باخ:
 کُيْتِيرِمک.
 کُيْتِيرِکْدِيرِيلِمک { کُيْتِيرِکْدِيرِيلِمَاک } باخ:
 کُيْتِيرِيلِمک.
 کُيْتِيرِکْمک { کُيْتِيرِکْمَاک } باخ: کُيْتِيرِکْمک.
 کُيْتِيرِکْمک { کُيْتِيرِکْمَاک } باخ: کُيْتِيرِکْمک.

Figure 11. Sample from Behzadi 2003, showing ARABIC LETTER FARSI YEH WITH INVERTED V.

۱- سو گتيرميش چو کوننتو.
 ۲- (ماوايئ، موفه اله گلْميش).
 ۳- اَهک اَهنگ شيره سي، اَهنگو.
 ۴- از به تکان نخوردن سوسودان تکان
 يْتِمَه مک (تريمنه مک)، بيرششئين
 دوروموده يْتِمَه مک.
 ۵- از به نجنبيدن باخ به از به تکان
 نخوردن.
 ۶- از آسياب افتادن سس - سميرثا تماق،

Figure 12. Sample from Behzadi 2004, showing ARABIC LETTER FARSI YEH WITH INVERTED V.

کاتبان چهار حرف فارسی را هم به شکل: «پ، چ، ژ، گ» (ک، گ) و هم به صورت «ب، ج، ز، ک»، و کلمات «که، چه» را به صورت های «کی، که گ، چی، چه، چ (یا: جی، جه، ج)» نوشته اند، قاعده «ذال فارسی» را مراعات کرده اند، حرف «که» را در بسیاری از موارد به کلمه بعد متصل نوشته اند، و در بعضی از نسخه ها «فاء اعجمی» (= و) را در برخی از موارد رعایت نموده اند. این شیوه

Figure 13. Sample from Matini 1968, showing ARABIC LETTER KEHEH WITH TWO DOTS ABOVE and ARABIC LETTER KEHEH WITH THREE DOTS BELOW.

اگرچه «کیسو» به معنی موی جعد در فرهنگها یافت نمی شود، اما به گمان من حتماً رواج داشته است. زیرا در فرهنگها نام دارویی «کیسو» آمده است که به عربی «جعه» خوانده می شود، پس اگر کیسو به کار برده نمی شده، جعه به کیسو ترجمه می شد و چون آن را کیسو می خوانند، معلوم است که کیسو به مفهوم کیسو حتماً رایج بوده است. در زبانهای هندی نیز کلمه «کیس» [به مفهوم موی جعد] رایج است.

۳. بالای حرف گاف معمولاً دو نقطه گذارده شده است. این دو نقطه همیشه بر مرکز آن قرار دارد. نمونه: گفت، گردن، گیرد، دیگر، بگرفت، گاو، چکونه، سنگی، بزگ. هرگاه گاف کوفی که مرکزش به دور خود پیچیده است،

Figure 14. Sample from Ravaghi 2005, showing ARABIC LETTER KEHEH WITH TWO DOTS ABOVE.

از کینه تناء سمین بشنیت او بوم سالها بوم بر کیان آوره بار
بهر راهی که او یان ادشی اجان راه روخیاران بدامان دل شه جی

این دو بیت با هم ارتباطی ندارند و این نکته از قافیه های آنها پیداست. به علاوه از وزن مختل آنها معلوم می شود که بعضی از کلمات آنها تصحیف شده است. معنی هیچ یک از آنها نیز روشن نیست و تنها مصراع سوم و بعضی از کلمات سایر مصراعها مشخص است که اینک آنها را توضیح می دهیم.

سه نقطه در زیر حرف اول کلمه کنه نشان می دهد که این حرف باید گاف g تلفظ شود.

Figure 15. Sample from Sadeghi 2003, showing ARABIC LETTER KEHEH WITH THREE DOTS BELOW and ARABIC LETTER YEH WITH TWO DOTS ABOVE.

TABLE XXX - Row 06: ARABIC

	060	061	062	063	064	065	066	067
0	ﻝ	ﻝ̣		ذ	-	ﻝ̣	ﺀ	ﻝ̣
1	ﻝ̣	ﻝ̣	ﺀ	ر	ف	ﻝ̣	ﺀ	أ
2	ﻝ̣	ﻝ̣	ﺀ	ز	ق	ﻝ̣	ﺀ	ﺀ
3	ﻝ̣	ﻝ̣	ﺀ	س	ك	ﻝ̣	ﺀ	ﻝ̣
4		ﻝ̣	ﺀ	ش	ل	ﻝ̣	ﺀ	ﺀ
5		ﻝ̣	ﺀ	ص	م	ﻝ̣	ﺀ	ﺀ
6	ﻝ̣	ﻝ̣	ﺀ	ض	ن	ﻝ̣	ﺀ	ﺀ
7	ﻝ̣		ﺀ	ط	ه	ﻝ̣	ﺀ	و
8	ﻝ̣		ﺀ	ظ	و	ﻝ̣	ﺀ	ﺀ
9	ﻝ̣		ﺀ	ع	ى	ﻝ̣	ﺀ	ﺀ
A	ﻝ̣		ﺀ	غ	ي	ﻝ̣	ﺀ	ﺀ
B	ﻝ̣	ﺀ	ﺀ	ﺀ	ﻝ̣	ﻝ̣	ﺀ	ﺀ
C	ﺀ		ﺀ	ﻝ̣	ﻝ̣	ﻝ̣	ﺀ	ﺀ
D	ﺀ		ﺀ	ﺀ	ﻝ̣	ﻝ̣	ﺀ	ﺀ
E	ﻝ̣	ﺀ	ﺀ	ﻝ̣	ﻝ̣	ﻝ̣	ﺀ	ﺀ
F	ﻝ̣	ﺀ	ﺀ	ﺀ	ﻝ̣		ﺀ	ﺀ

TABLE XXX - Row 06: ARABIC

hex	Name
00	ARABIC NUMBER SIGN
01	ARABIC SIGN SANAH
02	ARABIC FOOTNOTE MARKER
03	ARABIC SIGN SAFHA
04	(This position shall not be used)
05	(This position shall not be used)
06	<i>ARABIC-INDIC CUBE ROOT</i>
07	<i>ARABIC-INDIC FOURTH ROOT</i>
08	<i>ARABIC RAY</i>
09	<i>ARABIC-INDIC PER MILLE SIGN</i>
0A	<i>ARABIC-INDIC PER TEN THOUSAND SIGN</i>
0B	AFGHANI SIGN
0C	ARABIC COMMA
0D	ARABIC DATE SEPARATOR
0E	ARABIC POETIC VERSE SIGN
0F	ARABIC SIGN MISRA
10	ARABIC SIGN SALLALLAHOU ALAYHE WASSALLAM
11	ARABIC SIGN ALAYHE ASSALLAM
12	ARABIC SIGN RAHMATULLAH ALAYHE
13	ARABIC SIGN RADI ALLAHOU ANHU
14	ARABIC SIGN TAKHALLUS
15	ARABIC SMALL HIGH TAH
16	<u>ARABIC SMALL HIGH ALEF-LAM-YEH</u>
17	(This position shall not be used)
18	(This position shall not be used)
19	(This position shall not be used)
1A	(This position shall not be used)
1B	ARABIC SEMICOLON
1C	(This position shall not be used)
1D	(This position shall not be used)
1E	ARABIC TRIPLE DOT PUNCTUATION MARK
1F	ARABIC QUESTION MARK
20	(This position shall not be used)
21	ARABIC LETTER HAMZA
22	ARABIC LETTER ALEF WITH MADDA ABOVE
23	ARABIC LETTER ALEF WITH HAMZA ABOVE
24	ARABIC LETTER WAW WITH HAMZA ABOVE
25	ARABIC LETTER ALEF WITH HAMZA BELOW
26	ARABIC LETTER YEH WITH HAMZA ABOVE
27	ARABIC LETTER ALEF
28	ARABIC LETTER BEH
29	ARABIC LETTER TEH MARBUTA
2A	ARABIC LETTER TEH
2B	ARABIC LETTER THEH
2C	ARABIC LETTER JEEM
2D	ARABIC LETTER HAH
2E	ARABIC LETTER KHAH
2F	ARABIC LETTER DAL
30	ARABIC LETTER THAL
31	ARABIC LETTER REH
32	ARABIC LETTER ZAIN
33	ARABIC LETTER SEEN
34	ARABIC LETTER SHEEN
35	ARABIC LETTER SAD
36	ARABIC LETTER DAD
37	ARABIC LETTER TAH
38	ARABIC LETTER ZAH
39	ARABIC LETTER AIN
3A	ARABIC LETTER GHAIN
3B	ARABIC LETTER SEEN WITH INVERTED V
3C	<u>ARABIC LETTER KAF WITH TWO DOTS ABOVE</u>
3D	<u>ARABIC LETTER KEHEH WITH TWO DOTS ABOVE</u>
3E	<u>ARABIC LETTER KEHEH WITH THREE DOTS BELOW</u>
3F	<u>ARABIC LETTER YEH WITH INVERTED V</u>
40	ARABIC TATWEEL
41	ARABIC LETTER FEH
42	ARABIC LETTER QAF
43	ARABIC LETTER KAF
44	ARABIC LETTER LAM
45	ARABIC LETTER MEEM
46	ARABIC LETTER NOON
47	ARABIC LETTER HEH
48	ARABIC LETTER WAW
49	ARABIC LETTER ALEF MAKSURA
4A	ARABIC LETTER YEH
4B	ARABIC FATHATAN
4C	ARABIC DAMMATAN
4D	ARABIC KASRATAN
4E	ARABIC FATHA
4F	ARABIC DAMMA
50	ARABIC KASRA
51	ARABIC SHADDA
52	ARABIC SUKUN
53	ARABIC MADDAAH ABOVE
54	ARABIC HAMZA ABOVE
55	ARABIC HAMZA BELOW
56	ARABIC SUBSCRIPT ALEF
57	ARABIC INVERTED DAMMA
58	ARABIC MARK NOON GHUNNA

hex	Name
59	ARABIC ZWARAKAY
5A	ARABIC VOWEL SIGN SMALL V ABOVE
5B	ARABIC VOWEL SIGN INVERTED SMALL V ABOVE
5C	ARABIC VOWEL SIGN DOT BELOW
5D	ARABIC REVERSED DAMMA
5E	ARABIC FATHA WITH TWO DOTS
5F	(This position shall not be used)
60	ARABIC-INDIC DIGIT ZERO
61	ARABIC-INDIC DIGIT ONE
62	ARABIC-INDIC DIGIT TWO
63	ARABIC-INDIC DIGIT THREE
64	ARABIC-INDIC DIGIT FOUR
65	ARABIC-INDIC DIGIT FIVE
66	ARABIC-INDIC DIGIT SIX
67	ARABIC-INDIC DIGIT SEVEN
68	ARABIC-INDIC DIGIT EIGHT
69	ARABIC-INDIC DIGIT NINE
6A	ARABIC PERCENT SIGN
6B	ARABIC DECIMAL SEPARATOR
6C	ARABIC THOUSANDS SEPARATOR
6D	ARABIC FIVE POINTED STAR
6E	ARABIC LETTER DOTLESS BEH
6F	ARABIC LETTER DOTLESS QAF
70	ARABIC LETTER SUPERScript ALEF
71	ARABIC LETTER ALEF WASLA
72	ARABIC LETTER ALEF WITH WAVY HAMZA ABOVE
73	ARABIC LETTER ALEF WITH WAVY HAMZA BELOW
74	ARABIC LETTER HIGH HAMZA
75	ARABIC LETTER HIGH HAMZA ALEF
76	ARABIC LETTER HIGH HAMZA WAW
77	ARABIC LETTER U WITH HAMZA ABOVE
78	ARABIC LETTER HIGH HAMZA YEH
79	ARABIC LETTER TTEH
7A	ARABIC LETTER TTEHEH
7B	ARABIC LETTER BEEH
7C	ARABIC LETTER TEH WITH RING
7D	ARABIC LETTER TEH WITH THREE DOTS ABOVE DOWNWARDS
7E	ARABIC LETTER PEH
7F	ARABIC LETTER TEHEH

TABLE XXX - Row 07: ARABIC SUPPLEMENT

	075	076	077
0	پ	ف	ٹ
1	ث	ف	ڑ
2	پ	ن	ط
3	ت	ش	آ
4	ن	ک	آ
5	ب	م	ی
6	ن	م	ی
7	ح	ن	ی
8	ج	ن	و
9	ب	ن	و
A	د	ل	ے
B	ر	ز	ے
C	ش	ڑ	ج
D	ع	ش	س
E	ع	ح	ت
F	غ	ح	ث

G = 00
P = 00

TABLE XXX - Row 07: ARABIC SUPPLEMENT

hex	Name
50	ARABIC LETTER BEH WITH THREE DOTS HORIZONTALLY BELOW
51	ARABIC LETTER BEH WITH DOT BELOW AND THREE DOTS ABOVE
52	ARABIC LETTER BEH WITH THREE DOTS POINTING UPWARDS BELOW
53	ARABIC LETTER BEH WITH THREE DOTS POINTING UPWARDS BELOW AND TWO DOTS ABOVE
54	ARABIC LETTER BEH WITH TWO DOTS BELOW AND DOT ABOVE
55	ARABIC LETTER BEH WITH INVERTED SMALL V BELOW
56	ARABIC LETTER BEH WITH SMALL V
57	ARABIC LETTER HAH WITH TWO DOTS ABOVE
58	ARABIC LETTER HAH WITH THREE DOTS POINTING UPWARDS BELOW
59	ARABIC LETTER DAL WITH TWO DOTS VERTICALLY BELOW AND SMALL TAH
5A	ARABIC LETTER DAL WITH INVERTED SMALL V BELOW
5B	ARABIC LETTER REH WITH STROKE
5C	ARABIC LETTER SEEN WITH FOUR DOTS ABOVE
5D	ARABIC LETTER AIN WITH TWO DOTS ABOVE
5E	ARABIC LETTER AIN WITH THREE DOTS POINTING DOWNWARDS ABOVE
5F	ARABIC LETTER AIN WITH TWO DOTS VERTICALLY ABOVE
60	ARABIC LETTER FEH WITH TWO DOTS BELOW
61	ARABIC LETTER FEH WITH THREE DOTS POINTING UPWARDS BELOW
62	ARABIC LETTER KEHEH WITH DOT ABOVE
63	ARABIC LETTER KEHEH WITH THREE DOTS ABOVE
64	ARABIC LETTER KEHEH WITH THREE DOTS POINTING UPWARDS BELOW
65	ARABIC LETTER MEEM WITH DOT ABOVE
66	ARABIC LETTER MEEM WITH DOT BELOW
67	ARABIC LETTER NOON WITH TWO DOTS BELOW
68	ARABIC LETTER NOON WITH SMALL TAH
69	ARABIC LETTER NOON WITH SMALL V
6A	ARABIC LETTER LAM WITH BAR
6B	ARABIC LETTER REH WITH TWO DOTS VERTICALLY ABOVE
6C	ARABIC LETTER REH WITH HAMZA ABOVE
6D	ARABIC LETTER SEEN WITH TWO DOTS VERTICALLY ABOVE
6E	ARABIC LETTER HAH WITH SMALL ARABIC LETTER TAH BELOW
6F	ARABIC LETTER HAH WITH SMALL ARABIC LETTER TAH AND TWO DOTS
70	ARABIC LETTER SEEN WITH SMALL ARABIC LETTER TAH AND TWO DOTS
71	ARABIC LETTER REH WITH SMALL ARABIC LETTER TAH AND TWO DOTS
72	ARABIC LETTER HAH WITH SMALL ARABIC LETTER TAH ABOVE
73	ARABIC LETTER ALEF WITH EASTERN ARABIC-INDIC DIGIT TWO ABOVE
74	ARABIC LETTER ALEF WITH EASTERN ARABIC-INDIC DIGIT THREE ABOVE
75	ARABIC LETTER FARSI YEH WITH EASTERN ARABIC- INDIC DIGIT TWO ABOVE
76	ARABIC LETTER FARSI YEH WITH EASTERN ARABIC- INDIC DIGIT THREE ABOVE
77	ARABIC LETTER FARSI YEH WITH EASTERN ARABIC- INDIC DIGIT FOUR BELOW
78	ARABIC LETTER WAW WITH EASTERN ARABIC-INDIC DIGIT TWO ABOVE
79	ARABIC LETTER WAW WITH EASTERN ARABIC-INDIC DIGIT THREE ABOVE
7A	ARABIC LETTER YEH BARREE WITH EASTERN ARABIC- INDIC DIGIT TWO ABOVE
7B	ARABIC LETTER YEH BARREE WITH EASTERN ARABIC- INDIC DIGIT THREE ABOVE
7C	ARABIC LETTER HAH WITH EASTERN ARABIC-INDIC DIGIT FOUR BELOW
7D	ARABIC LETTER SEEN WITH EASTERN ARABIC-INDIC DIGIT FOUR ABOVE
7E	ARABIC LETTER FARSI YEH WITH TWO DOTS ABOVE
7F	ARABIC LETTER FARSI YEH WITH THREE DOTS ABOVE

hex	Name

A. Administrative

1. Title

Proposal to encode eight Arabic characters for Persian and Azerbaijani in the UCS

2. Requester's name

Michael Everson, Roozbeh Pournader, and Elnaz Sarbar.

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

4. Submission date

2006-10-24

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

No.

1b. Proposed name of script

1c. The proposal is for addition of character(s) to an existing block

Yes.

1d. Name of the existing block

Arabic and Arabic Supplement.

2. Number of characters in proposal

8.

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

Category A.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes.

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?

Amir Hedayaty (Sharif University of Technology, Tehran).

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Specialists in Persian linguistics and speakers of Azerbaijani.

4a. The context of use for the proposed characters (type of use; common or rare)

Common.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

Dictionaries, magazines, and linguistic publications.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Keep with other Arabic characters.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

No.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

Yes.

11b. If YES, is a rationale for such use provided?

No.

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?