

بررسی فونت‌های یونی‌کد رایانه‌ای فارسی و عربی

فریبا سلطانی‌نژاد*

چکیده

پس از ورود رایانه به کشور، در زمینه ایجاد الگوی چیدمان استاندارد حروف فارسی بر روی صفحه‌کلید و نیز ایجاد فونت‌های استاندارد، متولی معتبر و واحدی وجود نداشت، در نتیجه، شرکت‌ها و مؤسسات مختلف به دلیل نیاز بازار و خواسته‌های رو به گسترش جامعه، به ابداع خود، الگوی‌های مختلفی همانند استانداردهای غیررسمی ویندوز ۱۲۵۶، ایران‌سیستم، پانید، سایه و فارسی‌سازهای مختلفی را که اغلب متأثر از زبان و خط عربی بودند، به عنوان طرح حروف صفحه‌کلید و فونت ارائه کردند و از این نظر لطامات شدید و زیانباری به زبان و متون فارسی و نیز پایگاه‌های اطلاعات ملی و اداری و علمی وارد شد. به عنوان مثال، اگر گنجینه‌های ادب فارسی، همانند *شاهنامه*، با فونت‌های فارسی رایج و کدپیچ‌های مقتبس از عربی ذخیره شود و کسی بخواهد دنبال مصرعی شامل حرف «ی» بی نقطه فارسی بگردد، به دلیل وجود «ی» نقطه‌دار عربی هیچ موردی پیدا نمی‌کند، چراکه *شاهنامه* فارسی به سهو و ناخواسته به صورت عربی ذخیره شده است. در مقاله حاضر ابتدا درباره این تنگنا بحث، سپس پیشنهاد شده که راه حل گریز از آن، ایجاد، توسعه و استفاده از فونت‌هایی است که براساس یونی‌کد و استانداردهای مؤسسه تحقیقات صنعتی ایران تهیه شده‌اند، آنگاه با مقایسه فونت‌های رایج و برشمردن نقاط ضعف آنها نتیجه گرفته شده است که در صورت ادامه رویه فعلی، تمامی متون دیجیتال فارسی، اعم از گنجینه‌های ادبی، پایگاه‌های اطلاعاتی، متون علمی و... با بحران اسفناکی روبه‌رو خواهند بود.

کلیدواژه‌ها: فونت، متن دیجیتال، نمایش دیجیتال، متون، فونت‌های استاندارد، حروف فارسی و عربی، یونی‌کد، متون دیجیتال فارسی و عربی.

*. عضو هیئت علمی دانشگاه آزاد اسلامی - واحد جیرفت.

تاریخ دریافت: ۸۸/۱/۲۹، تاریخ پذیرش: ۸۸/۵/۱۲

fvs1234@yahoo.com

مقدمه

از چند سال پیش، در کشورمان، استفاده از رایانه با سرعت سرسام‌آوری جای خود را در تمامی عرصه‌ها باز کرد و سیل رایانه‌های شخصی و تجهیزات جانبی آنها به‌سوی کشور سرازیر شد. اما باید اعتراف کرد که با وجود اینکه سرعت سوق به‌سوی فناوری دیجیتال در ایران روند خوبی را طی کرده، در زمینهٔ ارائهٔ اطلاعات و پردازش آن به زبان فارسی تشتتی ایجاد شده است. یکی از عوامل مؤثر در این ناهماهنگی، نبود الگویی واحد برای ذخیره و پردازش و نمایش اطلاعات بر روی رسانه‌های جدید اطلاع‌رسانی مانند رایانه در سطح ملی است.

نرم‌افزارهای متفاوت، با فرمت‌های مختلف، کدهای فارسی گوناگون و... در حال استفاده‌اند و روزانه از اطلاعات زیادی را در خود جای می‌دهند. اگر از آن دسته از مراکزی که به‌دلیل نداشتن آگاهی کافی، اطلاعات را به‌صورت ناقص جمع‌آوری و وارد می‌کنند، بگذریم، به اختلاف و اعمال سلیقه‌های مختلف در سایر مراکز خواهیم رسید که برای نمونه به اختلاف درمورد کدنویس‌های به‌کاررفته برای حروف فارسی روی رایانه می‌توان اشاره کرد.

درمورد مراکزی که به هر حال مشغول سرمایه‌گذاری در بخش ورود، پردازش و نمایش اطلاعات هستند، مسئله به نوع دیگری خود را نشان خواهد داد. این‌گونه مراکز تا زمانی که پای خود را از محدودهٔ مرکز خود فراتر نگذاشته‌اند، مشکلی نخواهند داشت، ولی به‌محض اینکه بخواهند با مراکز اطلاعاتی و تحقیقاتی دیگر ارتباط برقرار کنند یا به مبادلهٔ اطلاعات با این مرکز بپردازند، متوجه خواهند شد که سال‌ها سرمایه‌های خود را برباد داده‌اند. همین مشکل در سطح ملی برای ایجاد یک مرکز اطلاعات ملی رخ خواهد نمود. زمانی این مشکل ملی بیشتر نمود پیدا می‌کند که بحث قرارگرفتن اطلاعات بر روی شبکهٔ جهانی اینترنت داغ‌تر شود. اینترنت به‌عنوان کلیدی برای ارتباط با دیگر مراکز اطلاعاتی به‌علت در دسترس بودن آسان و نیز حجم عظیم اطلاعات موجود در آن، یکی از مهم‌ترین موضوعاتی خواهد بود که

به‌علت نبود یک سیستم جهانی برای ذخیره، بازیابی، پردازش و نمایش اطلاعات و به‌طور کلی مبادله اطلاعات که جنبه ملی نیز داشته باشد، نقاط ضعفی دارد که ما را از بهره‌برداری مناسب در جهت منافعمان باز می‌دارد.

تعریف فونت

به مجموعه حروفی که با رابطه‌های گرافیکی، تصویر نوشتاری یک زبان را نشان می‌دهند، در زبان انگلیسی «فونت»^۱ گفته می‌شود. در زبان فارسی، واژه «قلم» را به‌جای «فونت» به‌کار می‌برند؛ اما در بین افرادی که با رایانه کار می‌کنند و یا با آن آشنایی دارند، بیشتر از کلمه انگلیسی «فونت» به‌جای «قلم» استفاده می‌شود. سائز (اندازه) یک فونت براساس فاصله بین پایین‌ترین نقطه کوتاه‌ترین حرف تا بالاترین نقطه بلندترین حرف محاسبه می‌شود. پس هنگامی که سائز یک قلم را ۱۰ تعریف می‌کنیم، یعنی فاصله بین دو نقطه مذکور ۱۰ نقطه است.

معرفی اولین فونت فارسی

شرکت کاتب، اولین شرکتی بود که توانست قلم فارسی را در رایانه‌های IBM^۲ و سازگار با آن طراحی کند. این شرکت، پایه‌گذار اساس حروف‌چینی فارسی دیجیتال در ایران و جهان بود.

آغاز حروف‌چینی فارسی با رایانه

در سال ۱۳۶۳ (مطابق با ۱۹۸۴ میلادی)، شرکت کاتب (مفیدرایانه) موفق شد با استفاده از فناوری روز و چاپگرهای لیزری شرکت HP، حروف فارسی را چاپ کند. این اولین بار بود که شرکتی در سطح جهان توانسته بود فونت فارسی را طراحی کند و بر روی چاپگر لیزری

1. font

۲. International Business Machines: ماشین‌های تجاری بین‌المللی.

چاپ کند؛ و این شروع تولید نرم افزارهای فارسی بود. از حسن فیروزخانی می توان به عنوان پدر نشر رومیزی فارسی نام برد. وی با سرمایه گذاری بسیار بالایی که در آن زمان انجام داد و با تلاش بی وقفه توانست این صنعت دیرپا را به خود مدیون کند.

اولین نرم افزار حروف چینی

اولین نرم افزار حروف چینی را شرکتی امریکایی برای ایرانیان نوشته بود که در محیط dos ver 1.5 کار می کرد. این نرم افزار، این قابلیت را به شما می داد که متن را بر روی چاپگرهای ماتریس نقطه ای^۱ چاپ کنید. البته ابتدا باید بر روی IC^۲ (مدار مجتمع) آن، تنها فونت فارسی که باز هم آن شرکت امریکایی آن را طراحی کرده بود و ۸ بیت بود و فقط از ۱۲۸ حرف استفاده می کرد، با سخت افزارهایی خاص کپی می شد. اولین رایانه ای که روی آن حروف چینی فارسی انجام شد، رایانه ای با cpu مدل ۸۰۸۶ با ۱۲۸ کیلوبایت ram بود که فقط یک فلاپی درایو پنج و یک چهارم اینچ و ظرفیتی معادل ۳۶۰ کیلوبایت داشت. این رایانه، ساخت شرکت IBM بود.

تاریخچه نرم افزارها و واژه پردازهای فارسی و عربی در ایران

نرم افزار نقش

نرم افزار نقش را که ابتدا با نام Wordp نوشته شده بود، مهندس پورنقشبند - از فارغ التحصیلان دانشگاه صنعتی شریف - نخست در شرکت کاتب نوشت؛ ولی بعدها آن را بسیار تکمیل کرد و با نام «نقش» روانه بازار ساخت. وی اولین فردی بود که توانسته بود اعراب را به صورت مستقل بر روی حروف فارسی اضافه کند. در این نرم افزار، هر حرف قابلیت داشت هر نوع اعرابی را رو یا زیر خود قرار دهد.

1. DotMatrix

2. Integrated Circuit

نرم افزار پیشکار

این نرم افزار در شرکت سینا نوشته شده بود؛ در محیط داس کار می کرد و قابلیت داشت قبل از چاپ، پیش نمایشی در اختیار کاربر قرار دهد. البته در آن زمان، این پیش نمایش به صورت 'Wysiwyg' نبود.

نرم افزار ام ال اس (MLS)

نرم افزار ام ال اس در یک شرکت خارجی نوشته شده بود؛ بر روی داس کار می کرد اما محیط آن کاملاً گرافیکی بود و اولین نرم افزاری بود که در ایران، سیستم Wysiwyg داشت و به همین جهت در آن برهه زمانی توانست مشتریان زیادی را به خود جلب کند. اما به دلیل فارسی نبودن کامل و نداشتن خدمات پس از فروش، جای خود را به زرنگار داد.

نرم افزار زرنگار

انقلاب حروف چینی فارسی رایانه ای زمانی رخ داد که شرکت سینا محصول خود به نام زرنگار را به ایرانیان معرفی کرد. این نرم افزار به طور کامل فارسی بود. خدمات پس از فروش قوی داش، و بسیار پرسرعت بود؛ و جالب اینکه دو حالت پیش نمایش داشت: یکی به صورت ساده، دیگری به صورت پیش نمایش قبل از چاپ که کاربر می توانست کاملاً صفحه چاپی خود را روی مونیتور ببیند. زرنگار از آن پس تا حدود ده سال هیچ رقیبی نداشت. تا اینکه ورد (word) توانست جای کلیه نرم افزارهای نشر رومیزی را در رایانه برای خود اشغال کند.

ویندوز فارسی سینا

اولین ویندوز فارسی را شرکت سینا ساخت. این شرکت با تغییراتی که در ویندوز ۳،۱

۱. What You See Is What You Get: آنچه را که می بینید، همان است که دریافت کرده اید.

عربی داده بود، توانست چهار حرف فارسی را به طریقی خاص اضافه کند؛ همچنین حدود شانزده فونت فارسی را طراحی و به این ساختار از ویندوز اضافه کند. اولین فونت فارسی را نیز شرکت سینا بر روی همین ساختار از ویندوز طراحی و نصب کرده بود.

Word فارسی

ورد عربی نسخه شش را چندین شرکت از جمله سینا فارسی کردند که هم بر روی ویندوز ۳,۱ و هم بر روی ویندوز ۹۵ فارسی کار می کرد. ورد از آن زمان به سبب پایین بودن قیمت آن در ایران و ساده بودن یادگیری آن، چنان جایی در دل ها باز کرد که در حال حاضر می توان گفت نرم افزار استاندارد نشر رومیزی در ایران شناخته می شود. اما بسیاری از کاربران نمی دانند که این نرم افزار - چنان که از نام آن هم مشخص است: MS Office Word - یک نرم افزار کاملاً اداری است و برای کتابهایی که به عنوان مثال پاورقی های سنگین داشته باشند، هیچ کاربردی ندارد و جز دردسرهای اضافی برای حروف چین و صفحه آرا، کار دیگری انجام نمی دهد. بسیار دیده شده است که در انتها مجبور شده اند کار را با نرم افزار زرنگار از ابتدا حروف چینی و صفحه آرای کنند.

ویندوز ۹۵ تا xp فارسی

هنگامی که ویندوز ۹۵ عربی روانه بازار شد، چندین شرکت از جمله سینا این ویندوز را فارسی کردند و آن را روانه بازار ساختند. این کار (یعنی ایجاد چند نمونه فارسی از ویندوز عربی) به خاطر نداشتن تجربه کافی باعث شد مشکلات بسیاری در زمینه نشر رومیزی و حتی حروف چینی فارسی در ویندوز پیش بیاید که هنوز هم بعضی کاربران با آن مشکلات مواجهند:

الف - فونت‌های فارسی برروی ویندوز ۹۵:

از همان ابتدا شرکت‌ها سعی کردند نوعی «کدپیج» (Code Page) را برای زبان فارسی ایجاد کنند و آن را سوای کدپیج عربی به بازار ارائه دهند؛ اما این کار به‌هیچ‌وجه پیش نرفت، چراکه هیچ تعاملی بین این مؤسسات و میکروسافت وجود نداشت و فقط باعث شد مشکلات بیشتری برای کاربران به‌وجود بیاید.

ب - فونت فارسی برروی ویندوز ۹۸:

با تکمیل‌تر شدن ویندوز و در نتیجه ویندوز ۹۸ عربی، ویندوز ۹۸ فارسی نیز به بازار آمد و اما این‌بار همه تصمیم گرفتند که با همان کدپیج ۱۲۵۶ میکروسافت کار کنند و فقط تغییراتی جزئی روی چند فایل اصلی ویندوز بدهند تا چهار حرف اصلی فارسی برروی ویندوز ۹۸ عربی فعال شود. پایه و اساس ویندوز ۹۸ ام‌ای نیز مانند ۹۸ بود و به همین دلیل بعد از آمدن نسخه عربی آن به بازار به‌سرعت فارسی شد. بین ویندوز ۹۸ و ام‌ای (نسخه‌های عربی) دیگر هیچ مشکلی برای نقل و انتقال فایل‌های حروف‌چینی‌شده وجود نداشت، چرا که دیگر استاندارد ۱۲۵۶ برای همه مؤسسات ایرانی استاندارد شده بود.

ج - فونت فارسی برروی ویندوز ۲۰۰۰ و XP:

فونت فارسی به هیچ عنوان بعد از ویندوز ۹۸ تغییرات اساسی نکرد و تا به حال همان فونت‌ها بین کاربران منتقل می‌شود و برروی ویندوز ۲۰۰۰ و XP نیز کار می‌کنند. اما بسیاری از کسانی که درباره سیستم طراحی و ساخت فونت اطلاعات جزئی داشتند، سعی کردند نام فونت را به نام شرکت خود و یا نام خود تغییر دهند؛ و همین امر، باعث بروز خطاهای بسیاری در ساختمان فونت می‌شد. متأسفانه تا به حال، هیچ شرکتی فونت‌های صحیح را به کاربران معرفی نکرده است و کاربران نمی‌دانند که بسیاری از مشکلاتشان

به دلیل استفاده از همین فونت‌های فارسی است.

نرم افزارهای فعلی با قابلیت حروف چینی فارسی

در حال حاضر بیشتر نرم افزارهای شرکت مایکروسافت به صورت چندزبانه کار می کنند و زبان های فارسی و عربی را هم پشتیبانی می کنند. نرم افزارهای گرافیکی^۱، نرم افزارهایی هستند که به کاربران فارس، افغان، عرب، و کلاً تمامی کاربرانی که به سیستم راست به چپ احتیاج دارند، خدمات می دهند.

مشکلات لیتوگرافی ها

آن قدر فیلم و زینگ و کاغذ و مرکب به هدر رفت - و هنوز هم می رود - البته کمتر از سابق) - تا دریافتند که باید از نرم افزارهایی استفاده کنند که حروف فارسی را به درستی پشتیبانی کنند. متأسفانه هنوز هم در بسیاری از روزنامه ها، و حتی روزنامه های معتبر، از بسیاری از نرم افزارهای غیراستاندارد فارسی شده، مانند پیچ میکر انگلیسی - که با پلاگین هایی فارسی شده است - استفاده می شود. جالب اینجا است که هرچند وقت یک بار در روزنامه های معتبر ایرانی می بینید که مثلاً یک صفحه با حروفی مانند حروف یونانی و درهم و برهم چاپ شده است؛ و هنوز هم همان مسائل سابق که «ببخشید آقای لیتوگراف! فونت ها را یادم رفته بیاورم»، «ببخشید فونت ها را کرو نکرده ام» و... در کوچه پس کوچه های نشر ایران به گوش می رسد.

مشکلات حروف چین ها

حروف چین ها نیز به چند دلیل سرگردان شده اند:

۱. انواع صفحه کلیدها با استانداردهای مختلف (البته فقط یکی استاندارد است)؛

۱. نرم افزارهایی مانند InDesign، PhotoShop و CorelDraw.

۲. یک فونت با چند اسم، حتی روی یک رایانه؛

۳. متغیربودن یک سایز مطلق در رایانه‌های مؤسسات مختلف.

برای نمونه، درمورد فونتی که چندین سال پیش، فردی به نام لوتوس آن را نامگذاری کرد، به نام‌های زیر برمی‌خوریم:

B-Lotus
Lotus
Lotoos
Lotous
Mazar-Lotous
Lotous-Mazar
IDLotus
MELotus

ریشه مشکلات در چیست؟

از زمانی که اولین گزارش «زبان فارسی و کامپیوتر» در سال ۱۳۵۶ در دانشکده ریاضی و کامپیوتر دانشگاه صنعتی شریف ارائه شد، تا امروز که شبکه اینترنت چهره دیگری به اطلاع‌رسانی داده است، مدت زیادی می‌گذرد. امروزه دیگر محدودیت‌های سخت‌افزاری یا نرم‌افزاری نمی‌تواند مانع پیاده‌سازی یک سیستم ذخیره‌سازی، نمایش، و تبادل اطلاعات چندزبانه شود. امروز، مؤسسات بزرگ استانداردسازی مانند ایزو (ISO) و Consortium 3W نیز در استانداردهایشان مشکلات و مسائل مربوط به جهانی‌سازی را در نظر می‌گیرند تا امر تبادل اطلاعات چندزبانه را تسهیل کنند. اما به نظر می‌رسد که به دلیل عدم تصور ایرانیان و فارسی‌زبان‌ها در این روند، زبان فارسی قدری غریب مانده و کمتر به آن توجه شده است. به عنوان مثال، هنوز در بین صدها مجموعه نویسه ثبت‌شده در اینترنت توسط IANA^۱، فقط یک مجموعه نویسه ثبت‌شده متعلق به زبان

۱. Internet Assigned Number Authority: نهاد شماره‌های اختصاصی اینترنت.

فارسی است که آن هم «کد پیج» اختصاصی شرکت آیبیم است. حتی درمورد استاندارد کلی تبادل اطلاعات نیز قالبی که مورد توافق همه باشد، وجود ندارد. سه قالب موجود - ایران سیستم، استاندارد ۲۹۰۰ و استاندارد ۳۳۴۲ - هریک ایراداتی دارند که سبب شده است شرکت‌ها و مؤسسات داخلی به جدول‌های خاص خود روی آورند تا بتوانند نیازهای خود را تاحدی رفع سازند. اخیراً راه‌حل‌هایی در هر یک از مسائل خاص مربوط به تبادل اطلاعات برای بین‌المللی‌سازی در نظر گرفته شده است که با وجود اینکه این موارد کامل‌تر از جداولی است که در ایران برای حل مشکلات تبادل اطلاعات زبان فارسی ایجاد شده، به دلیل نبود مراجع موثق درمورد خط و زبان فارسی برای استانداردگذاران، مسائل خاص این زبان یا در نظر گرفته نشده و یا به شکل ناقص منظور شده است. خوشبختانه در بسیاری از این استانداردها امکان گسترش بعدی در نظر گرفته شده است که روند تصحیح را تسهیل می‌کند.

معرفی یونی‌کد

از جمله استانداردهای بین‌المللی که کامل‌تر از بقیه استانداردهای موجود، نیازهای مربوط به تبادل اطلاعات چندزبانه را رفع کرده است، می‌توان به استاندارد یونی‌کد اشاره کرد. این استاندارد، که تقریباً تمامی شرکت‌های بین‌المللی رایانه‌ای، مانند آیبیم، مایکروسافت، و سان، و نیز مؤسسات ملی استاندارد در کشورهای مختلف جهان برای تبادل اطلاعات چندزبانه روی آن توافق کرده‌اند، و سرعت رشد بسیار زیادی در میان کاربران دارد. همچنین، در حال حاضر، کلیه استانداردهای جدیدی که برای شبکه اینترنت طراحی می‌شوند، دو استاندارد را به عنوان کدپیچ پیش فرض می‌پذیرند: استاندارد XML و زبان جاوا. به زبان ساده می‌توان گفت که یونی‌کد روشی برای تبدیل متون به رشته‌های عددی قابل ذخیره در رایانه است. روش‌های گوناگونی برای این کار وجود دارد؛ ولی مزیت یونی‌کد بر آنها

این است که روشی کاملاً جهانی است، به این معنی که حروف همه زبان‌های دنیا و تمامی علائم مورد استفاده همه مردم جهان در آن وجود دارد، در همه جا قابل نمایش است و به امکانات خاصی نیاز ندارد. البته یونی‌کد هنوز جوان است ولی امروزه بسیاری از نرم‌افزارهای رایج در جهان از جمله همه مرورگرهای جدید اینترنت آن را پشتیبانی می‌کنند. از مهم‌ترین مزایایی که یونی‌کد برای زبان فارسی مثل بسیاری زبان‌های دیگر دارد، می‌توان به موارد زیر اشاره کرد:

۱. در نسخه استاندارد هر نرم‌افزاری که از این استاندارد پشتیبانی کند، می‌توان فارسی نوشت یا متون فارسی را خواند؛ و بدین ترتیب، دیگر نیازی به تأمین نسخه‌های خاص فارسی یا عربی نیست.

۲. برای خواندن متون فارسی که در شرکت خاصی نوشته شده است، به داشتن فونت خاص آن شرکت نیازی نیست و هر متن فارسی که با استاندارد یونی‌کد، کدگذاری شده باشد، با هر فونت یونی‌کدی قابل مشاهده است.

۳. امکان استفاده همزمان از زبان‌های فارسی و انگلیسی در آن وجود دارد.

۴. بدون استفاده از فونت‌های خاص، امکان استفاده از علائم خاص را فراهم می‌کند. به بیان دیگر، استاندارد یونی‌کد، استاندارد جهانی کدگذاری کاراکترها است که برای پردازش رایانه‌ای متون به کار می‌رود. این استاندارد، همان کاراکترها و کدهای استاندارد ISO/IEC 10646 را دارد و کاملاً با آن سازگار است. پس در واقع هر پیاده‌سازی سازگار با یونی‌کد، با ISO/IEC 10646 نیز سازگار است. با یونی‌کد، امکان کدگذاری همه کاراکترهای مورد استفاده در نوشتن زبان‌های دنیا فراهم آمده است. در این استاندارد، از کدگذاری ۱۶ بیتی استفاده می‌شود که برای بیش از ۶۵۰۰۰ نویسه (کاراکتر) جا فراهم می‌کند.

۱. International Electrotechnical Commission: کمیسیون بین‌المللی الکترونیکی.

۲. International Organization for Standardization: سازمان بین‌المللی استاندارد.

اگرچه ۶۵۰۰۰ نویسه برای کدگذاری اکثر نویسه‌هایی که در زبان‌های مهم دنیا استفاده می‌شود، کافی است، با این حال، یونی‌کد شیوه گسترشی به نام UTF-16^۱ فراهم کرده است که امکان اضافه کردن حدود یک میلیون نویسه دیگر را نیز می‌دهد. این دامنه برای کلیه نویسه‌های عالم، از جمله پوشش کامل همه خط‌های باستانی (مانند خط میخی) نیز کافی است. یونی‌کد برای کلیه نویسه‌های مورد استفاده در زبان‌های عمده دنیا کد تعیین کرده است. این استاندارد، به دلیل گسترده بودن فضای تخصیص نویسه در آن، بسیاری از نمادهای لازم برای حروف چینی را نیز دربر گرفته است.

از خط‌های مورد پشتیبانی این استاندارد می‌توان به لاتین (دربرگیرنده اکثر زبان‌های اروپایی)، سیریلیک (روسی، صربی)، یونانی، عربی (شامل عربی، فارسی، اردو، کردی)، عبری، هندی، ارمنی، آشوری، چینی، کاتاکانا و هیراگانا (ژاپنی)، و هانگول (کره‌ای) اشاره کرد. به علاوه، تعداد زیادی نماد ریاضی و فنی علائم نقطه‌گذاری، پیکان و علامت‌های متفرقه در این استاندارد وجود دارد. در این استاندارد، برای علامت‌های ترکیب‌شونده یا اعراب‌ها نیز کدهایی در نظر گرفته شده است که از جمله آنها علامت‌هایی مانند ؟ و مد هستند که در ترکیب حروف پایه، حروف تغییر یافته‌ای مانند ؟ را می‌سازند.

به طور کلی، بعضی از مشخصات یونی‌کد عبارت است از:

۱. یکسان‌سازی نویسه‌های مشترک در چند زبان مختلف؛ به عنوان مثال، به علت یکسانی «ع» در زبان عربی و فارسی، برای آن یک کد در نظر گرفته می‌شود.
۲. در استاندارد یونی‌کد، نویسه‌های فارسی در بلوک مربوط به خط عربی قرار دارند. این بلوک برای دربرگرفتن نویسه‌های زبان‌هایی که در آنها از خط عربی استفاده می‌شود - مثل فارسی، اردو، پشتو، سندی، و کردی - گسترش یافته است. این بلوک، نشانه‌های قرآنی از قبیل نشانه‌های سجده و پایان آیه، و علائم وقف را نیز دربر دارد.

۱. Unicode Transformation Format: فرمت تغییر یونی‌کد.

۳. در یونی‌کد، با وجود یکی‌سازی کدهای حروف مشترک، برای حروف فارسی که بار معنایی یا نمایشی متفاوت با حروف عربی دارند، نویسه‌های جداگانه در نظر گرفته شده است؛ یعنی کلیه حروف خاص فارسی (پ، چ، ژ، گ) و نیز «ک» و «ی» فارسی که با حروف مشابه در عربی تفاوت نمایشی دارند، مکان جداگانه‌ای به خود اختصاص داده‌اند. کلیه اعراب‌های متداول تصور دارند و میان شکل فارسی/اردو و عربی ارقام نیز، به علت شکل و رفتار متفاوت، تفاوت‌هایی منظور شده است. از طرف دیگر، علائم نقطه‌گذاری مانند نقطه و فاصله که شکل یکسانی در خط‌های لاتین و عربی دارند، کد یکسان دارند. علائمی مانند پُرانتز نیز، باتوجه به جهت متن، آینه‌ای می‌شوند، به‌طور مثال، نویسه ۰۰۲۸ نماینده پُرانتز باز است، و نه پُرانتز سمت چپ. یونی‌کد، اتصال مجازی و فاصله مجازی را نیز تحت نام‌های «اتصال با عرض صفر» و «بی‌اتصال با عرض صفر» به رسمیت می‌شناسد. بدین ترتیب، ملاحظه می‌شود که برای حل مشکلات موجود، و نیز رفتن به‌سوی یک استاندارد مقبول و همه‌جانبه، استاندارد یونی‌کد، روشی مناسب به نظر می‌رسد.

استانداردهای صفحه‌کلید فارسی براساس یونی‌کد

سال‌ها پس از ورود رایانه به کشور، در زمینه ایجاد استاندارد و الگوی چیدمان استاندارد حروف فارسی و نیز ایجاد فونت‌های استاندارد، هیچ مؤسسه و متولی واحد و معتبری هیچ اقدامی انجام نداد و نتیجه آن شد که شرکت‌ها و مؤسسات مختلف، به دلیل نیاز بازار و خواسته‌های رو به گسترش جامعه که در استفاده از رایانه به شدت شتاب زده بود، به ابداع خود، الگوی‌های مختلفی را به عنوان طرح حروف صفحه‌کلید و فونت ارائه کردند که می‌توان به استانداردهای غیررسمی ویندوز ۱۲۵۶، ایران سیستم، پانیز و سایه و فارسی‌سازهای مختلفی اشاره کرد که در ابتدای مقاله نام برخی از آنها ارائه شد. اما رشد این رویه با توجه به همه‌گیر شدن استفاده از رایانه و ضرورت مبادله اسناد دیجیتال و عدم همخوانی آنها موجب

شد که مؤسسه استاندارد و تحقیقات صنعتی کشور به عنوان تنها متولی استانداردسازی، با تشکیل کارگروه‌ها و تیم‌های مرکب از متخصصان و شرکت‌های معتبر و مؤثر در فناوری اطلاعات، به ابداع استانداردهای مختلف اقدام کند؛ که در زیر به دو استاندارد مهم در این زمینه اشاره شده است.

استاندارد ۹۱۴۷ ماتصا^۱ برای صفحه کلید

استاندارد ۹۱۴۷ یا همان ISIRI 9147^۲، جدیدترین استاندارد است که مؤسسه استاندارد و تحقیقات صنعتی ایران برای چیدمان حروف و علائم خط فارسی بر روی صفحه کلید رایانه منتشر کرده است. این استاندارد در سال ۱۳۸۶ انتشار یافته است؛ و در حال حاضر (مرداد ۱۳۸۸)، آخرین و جدیدترین استاندارد در این زمینه محسوب می‌شود. ماتصا ۹۱۴۷ جایگزین استاندارد قدیمی‌تر ماتصا ۲۹۰۱ (ISIRI 2901) شده و آن را باطل اعلام کرده است. البته استاندارد ۹۱۴۷ فرق خیلی زیادی با استاندارد ۲۹۰۱ ندارد و به اصطلاح Compatible Backward^۳ است.

کسانی که دستشان به ۲۹۰۱ عادت کرده است، می‌توانند امیدوار باشند که خیلی زود به ۹۱۴۷ هم عادت می‌کنند. متأسفانه این استاندارد برخلاف استاندارد ۶۲۱۹، هنوز در بین کاربران، به‌ویژه کاربران ویندوز، رواج خیلی زیادی پیدا نکرده است. علت این امر، تنبلی کاربران و بی‌توجهی مدیران IT کشور در همه سطوح است. متأسفانه بیشتر کاربران فکر می‌کنند آنچه در ویندوز برایشان مهیا است، همان استاندارد واقعی است؛ در نتیجه تصور می‌کنند ماتصا ۹۱۴۷ صرفاً یک ابداع جدید است که نباید آن را زیاد جدی گرفت. تعدادی از کاربران هم چون این صفحه کلید، صفحه کلید پیش فرض ویندوزهای ویستا، ایکس پی یا ۹۸

۱. مؤسسه استاندارد و تحقیقات صنعتی ایران.

۲. این استاندارد را مؤسسه استاندارد و تحقیقات صنعتی ایران (ماتصا) برای تعریف و یکسان‌سازی استفاده از حروف الفبای فارسی در محیط رایانه در سال ۱۳۸۱ تدوین کرده است.

۳. سازگار با حالت برگشت به عقب.

نیست و یا چون به لحاظ فنی با نصب آن مشکل دارند، از آن استفاده نمی‌کنند. ماتصا ۹۱۴۷ باتوجه به استاندارد ماتصا ۶۲۱۹، یعنی استاندارد «فناوری اطلاعات - تبادل و شیوه نمایش اطلاعات فارسی براساس یونی‌کد» است. در صورت استفاده از چیدمان‌های غیراستاندارد، کاراکترهای غیراستاندارد (خارج از استاندارد ۶۲۱۹) هم تولید می‌شود. حضور تعداد زیادی «ی» دو نقطه عربی و «ک» عربی گواه بر استفاده از صفحه‌کلیدهای غیراستاندارد رایج در ویندوز است. ماتصا ۹۱۴۷ علاوه بر اینکه جای هر کدام از کاراکترهای مجاز را بر روی صفحه‌کلید به وضوح مشخص و استاندارد کرده، چند کار مهم دیگر را هم انجام داده است: ۱. حروف صرفاً عربی و غیرمجاز در استاندارد ۶۲۱۹ نیز در چیدمان ۹۱۴۷ حضور دارند. این حروف در کار با متون غیراستاندارد و ناسازگار با ماتصا ۶۲۱۹ و تبدیل و جست‌وجو در متون غیراستاندارد و نیز تحریر متون عربی کاربرد فراوانی دارد.

۲. علائم پرکاربرد غیرفارسی، مثل «@»، «»، «» و... هم در آرایش صفحه‌کلید موجود هستند. بدون وجود این کاراکترها در استاندارد برای ورود آنها باید یک بار صفحه‌کلید به انگلیسی عوض و دوباره به فارسی برگردانده می‌شد.

۳. اعداد موجود در ردیف بالایی صفحه‌کلید، اعداد فارسی واقعی (نه انگلیسی و نه عربی) را وارد می‌کنند و کلیدهای عددی ماشین حسابی هم همان اعداد انگلیسی همیشگی را وارد می‌کنند. ممکن است خیلی از ویندوزها اعداد فارسی و انگلیسی را به ظاهر به یک شکل و آن هم به شکل فارسی نمایش دهند. این قضیه فقط ظاهری است و اصل اطلاعات به همان صورت نادرست انگلیسی ذخیره می‌شود. علت این رفتار ویندوز، وجود Regional Settings است.

۴. کاراکترهای ویژه کار با متون دوجته که در الگوریتم شماره ۹ یونی‌کد معرفی شده، در این چیدمان موجود است و به خوبی پشتیبانی می‌شوند. این کاراکترها در صورتی که کاربران با نقش آنها آشنا باشند، کاربرد فراوانی در حل مشکلات راست به چپ‌نویسی

۱. تنظیمات زبان منطقه‌ای در کنترل پانل.

متون فارسی در محیط رایانه و اینترنت دارند. مثلاً با کمک این سری کاراکترها می‌توان در صفحات مطلقاً چپ به راست (انگلیسی)، متون ترکیبی فارسی - انگلیسی را به راحتی و بدون هرگونه مشکل و درهم‌ریختگی وارد کرد. خوشبختانه ماتصا ۹۱۴۷ در بیشتر لینوکس‌ها رعایت و پشتیبانی می‌شود. برای استفاده از این استاندارد در لینوکس، به راهنماهای موجود مراجعه فرمایید؛ و برای به‌کارگیری این استاندارد در ویندوزهای مختلف از جمله ایکس‌پی، ویستا و ویندوز ۷، از درایور مخصوصی که در این زمینه تهیه شده است باید استفاده کرد.

استاندارد ماتصا ۶۲۱۹ (ISIRI 6219)

این استاندارد در اردیبهشت ۱۳۸۱، برای تعریف و یکسان‌سازی استفاده از حروف و الفبای فارسی در محیط رایانه، در مؤسسه استاندارد و تحقیقات صنعتی ایران (ماتصا) تدوین شده است. این استاندارد در حال حاضر (تیر ۱۳۸۸)، جدیدترین استاندارد در این زمینه و جایگزین کلیه استانداردهای قدیمی‌تر ماتصا ۳۳۴۲، ماتصا ۲۹۰۰ و استانداردهای غیررسمی ویندوز ۱۲۵۶، ایران‌سیستم، پانیز و سایه است. ماتصا ۶۲۱۹ کاملاً براساس یونی‌کد است و گروهی از افراد متخصص و مسلط در این زمینه آن را نگاشته‌اند. مسائل مطرح‌شده در ماتصا ۶۲۱۹، از روز تدوین آن تاکنون مقبولیت روزافزونی داشته و روز به روز در نرم‌افزارهای بیشتری رعایت شده است. راحت‌تر شدن جست‌وجو و مقایسه اطلاعات فارسی که در منابع مختلف ثبت شده است، رفع مشکل انواع «ی» در فونت‌های مایکروسافت و بقیه، تمایز قطعی بین انواع «ی»ها و انواع «ک»ها، رواج گسترده زبان فارسی در محیط لینوکس و... ازجمله نتایج این استاندارد است. از آنجا که در استاندارد یونی‌کد به‌ازای بسیاری از حروف مثل «ی»، «ک»، «الف» و بسیاری از اعداد و نشانه‌ها و لیگاتورها تعداد زیادی کاراکتر یونی‌کد شبیه هم وجود دارد، بسیاری از افراد و کاربران در استفاده از آنها دچار ابهام می‌شوند؛ برای مثال، مطابق توضیحات بلوک عربی یونی‌کد، حدود ده «ی» مختلف وجود دارد که استفاده از هر کدام فقط در استاندارد یک یا چند کشور

مجاز است.

مهم‌ترین موضوعی که ماتصا ۶۲۱۹ به آن توجه کرده، مشخص کردن کاراکترهای مجاز و غیرمجاز برای خط و زبان فارسی در ایران است؛ به‌عنوان نمونه، در این استاندارد فقط حرف «ی» فارسی با کد U+06CC و «ئ» همزه‌دار با کد U+0626 برای خط و زبان فارسی مجاز اعلام شده است. موضوع مهم بعدی که ماتصا ۶۲۱۹ به آن توجه داشته، نحوه نمایش حروف و کلمات فارسی از دیدگاه چپ و راست‌چینی (و نه شکل قلم) است. با این قواعد، مشکلات مرتبط با به‌هم‌ریختگی حروف و کلمات فارسی به‌ویژه به‌هنگام استفاده همزمان با حروف و کلمات لاتین حل می‌شود. ماتصا ۶۲۱۹، الگوریتم شماره ۹ یونی کد را که به الگوریتم دوجهته معروف است، به‌عنوان مرجع و بخشی از خود معرفی کرده است. با رعایت صحیح استاندارد ماتصا ۶۲۱۹:

۱. حرف «ی» بی‌نقطه نجسب عربی با کد U+0649 که متون فارسی را کاملاً به‌هم می‌ریزد، از متون فارسی حذف می‌شود. از به‌هم‌ریختگی‌های خیلی آزاردهنده ایجادشده توسط این نوع «ی» می‌توان به تکه‌تکه‌شدن کلمات حاوی «ی» در بسیاری از موبایل‌های امروزی و بسیاری از نرم‌افزارهای قدیمی ویندوز اشاره کرد.

۲. به‌جای اعداد لاتین و یا حتی اعداد عربی، از اعداد فارسی استفاده می‌شود. این اعداد در ویندوزهایی که Regional Settings آنها تغییر پیدا کرده است، به‌صورت ظاهراً فارسی نمایش داده می‌شوند؛ ولی وقتی به PDF تبدیل می‌شوند یا در محیط‌های دیگری از طریق وب دیده می‌شوند، به همان صورت غلط لاتین مشاهده می‌شوند.

۳. از درهم‌ریختگی متون ترکیبی فارسی و انگلیسی در اکثر نرم‌افزارها و محیط‌های «Platform» امروزی جلوگیری می‌شود؛ به‌عنوان مثال: نمایش برعکس پراتنرها، جابه‌جایی حروف نشانه‌ای مثل سمی‌کالن، نقطه و... حتی وقتی که کلمات کاملاً انگلیسی هستند و صرفاً در یک محیط دوزبانه (دوجهته یا Bidirectional) نمایش داده می‌شوند.

ماتصا ۶۲۱۹ این کار را با استفاده از کاراکترهای ویژه الگوریتم شماره ۹ مثل EMBEDDING RIGHT-TO-LEFT انجام می‌دهد.

۴. استفاده از علائم در خط و زبان فارسی یکسان می‌شود. برای مثال، در این استاندارد، استفاده از دابل کوتیشن و تک کوتیشن رایج در متون انگلیسی ممنوع و به جای آنها، کاراکترهای «» که شبیه دو علامت کوچک‌تر یا بزرگ‌تر به هم چسبیده هستند به عنوان «گیومه فارسی» معرفی شده است. همچنین در این استاندارد، کاراکترهای مشخصی برای ممیز فارسی، جداکننده هزارگان فارسی و... در نظر گرفته شده است.

حروف فارسی معرفی شده در یونی‌کد، همگی براساس بلوک عربی هستند. دلیل آن هم این است که ما هیچ بلوک یا بخش واحدی در یونی‌کد برای خط فارسی یا دیگر خط یا زبان‌های مبتنی بر عربی، مثل اردو و کردی، نداریم. برای انجام دادن بعضی امور اولیه کار با متون یونی‌کد و استاندارد ماتصا ۶۲۱۹، ابزار کوچکی به نام «فردیس» وجود دارد که استفاده از آن در بعضی مواقع می‌تواند مفید باشد. استاندارد ماتصا ۶۲۱۹، کاری به چیدمان حروف فارسی در صفحه کلید ندارد. ولی خود استاندارد ماتصا ۹۱۴۷ که چیدمان حروف و علائم فارسی بر روی صفحه کلید رایانه را معین می‌کند، براساس همین استاندارد ماتصا ۶۲۱۹ طراحی شده است. ماتصا ۶۲۱۹، مستقل از شکل نمایشی (Glyph) حروف است؛ به طور مثال، نمی‌گوید انتهای حرف «ف» چقدر باید به بالا کشیده شده باشد.

جدول مقایسه فونت‌های رایج فارسی با عربی

درمورد بسیاری از فونت‌های رایج، مشکلاتی وجود دارد، از جمله: پیاده‌سازی غلط «ی» و «ک» و «ه» عربی و بعضی کاراکترهای خاص مورد نیاز مثل ممیز، درصد و جداکننده هزارگان فارسی (عربی) و فارسی کردن غلط اعداد انگلیسی. برای بررسی این مشکلات، بخش اعظمی از فونت‌های رایج و غیررایج فارسی، از نظر سازگاری با برنامه‌هایی مثل

Reporting Services^۱ و دیگر برنامه‌هایی که در مجموعه فونت‌های آنها، انتظار می‌رود به غیر از کاراکترهای فارسی، کاراکترهای انگلیسی هم پیاده‌سازی شده باشد، بررسی و مقایسه شد. برای این کار، فونت‌های معروف «سری بی برنارایانه» که تقریباً در هر ویندوزی پیدا می‌شود، فونت‌های «فارسی وب»، فونت‌های طراحی شده در «انجمن کاربران ایرانی اپل مکینتاش» یعنی «ایرماگ»، فونت‌های عربی «مایکروسافت»، فونت «ایران نستعلیق شورای عالی اطلاع‌رسانی»، فونت «فارسی آزاد»، فونت «دجاوو اوبونتو»، فونت‌های «مؤسسه غیرانتفاعی sil»، فونت‌های یونی‌کد مایکروسافت مثل تاهوما و بعضی فونت‌های متفرقه مثل «نصف» را بررسی و نتیجه در جدول ۱ خلاصه شده است.

مبنای مقایسه و رتبه‌بندی فونت‌ها، سازگاری با استانداردهای ۶۲۱۹ و ۹۱۴۷ (که نسخه جدیدتر استاندارد ۲۹۰۱ است) و ابزار گزارش‌گیری مایکروسافت یعنی Reporting Services MS SQL بوده است.

این استانداردها، آخرین و معتبرترین استانداردهای موجود در زمینه فارسی‌نویسی در رایانه هستند. همان‌طور که در جدول مقایسه پیدا است، عمده مشکل بر سر کاراکترهای خاص فارسی (عربی) مثل ممیز، اعداد، جداکننده هزارتایی‌ها و علامت مخصوص ریال است. فونت‌هایی که ناسازگاری آنها با استانداردهای بالا، بیش از حد خاصی بوده، از جدول مقایسه حذف شده‌اند؛ یعنی اگر فونتی در این بررسی هست ولی در جدول ۱ نیست، به علت ناسازگاری بیش از حد آن بوده است مثل فونت‌های «سری بی» و فونت «فارسی آزاد». البته باتوجه به معیارهای استاندارد ۶۲۱۹ و ۹۱۴۷ و مشکلات مربوط به Reporting Services فقط ۴ فونت «ترافیک فارسی وب»، «تبریز ایرماگ»، «شهرزاد» و «لطیف» توانسته‌اند از این آزمون سربلند بیرون بیایند. دو فونت آخر در مؤسسه غیرانتفاعی sil تهیه شده‌اند.

۱. سرویس‌های گزارش‌گیری



جدول ۱. مقایسه فونت‌های رایج فارسی

نام رایج فونت	نام فونت در مجموعه MS Office	تولید کننده فونت	میزان پشتیبانی از استاندارد صفحه کلید ۹۱۴۷	آیا امکان تایپ انگلیسی با همان فونت وجود دارد؟	آیا الفبا و حروف انگلیسی در خود فونت پیاده سازی شده است؟	MS SQL با Reporting Services 2008 سازگار است؟
۱. نازلی	Nazli	فارسی وب	به جز کاراکترهای : U+FD3F و U+FD3E	خیر	بله	بله
۲. ترافیک	Terafik	فارسی وب	به طور کامل	خیر	بله	بله
۳. رویا	Roya	فارسی وب	به طور کامل	خیر	بله	خیر
۴. کودک	Koodak	فارسی وب	به طور کامل	خیر	بله	خیر
۵. هما	Homa	فارسی وب	به طور کامل	خیر	بله	خیر
۶. تیترا	Titra	فارسی وب	به طور کامل	خیر	بله	خیر
۷. زر	XB Zar	ایرماگ	همه به جز کاراکترهای : U+066C و U+FD3E	بله	بله	بله
۸. نیلوفر	XB Niloofer	ایرماگ	همه به جز : U+066C که به شکل اشتباه پیاده سازی شده	بله	بله	بله
۹. ریاض	XB Riyaz	ایرماگ	همه به جز : U+066C	بله	بله	بله
۱۰. رویا	XB Roya	ایرماگ	همه به جز : U+066C	بله	بله	بله
۱۱. شفیق	XB Shafiq	ایرماگ	همه به جز : U+066C	بله	بله	بله
۱۲. شیراز	XB Shiraz	ایرماگ	همه به جز : U+066C	بله	بله	بله
۱۳. ثلث	XB Sols	ایرماگ	همه به جز : U+066C	بله	بله	بله
۱۴. تبریز	XB Tabriz	ایرماگ	به طور کامل	بله	بله	بله
۱۵. تیترا	XB Titre	ایرماگ	همه به جز : U+066C	بله	بله	بله
۱۶. تیترا سایه دار	XB Titre Shadow	ایرماگ	همه به جز : U+066C	بله	بله	بله
۱۷. یاقوت	XB Yagut	ایرماگ	همه به جز : U+066C	بله	بله	بله
۱۸. یاس	XB Yas	ایرماگ	همه به جز : U+066C	بله	بله	بله
۱۹. ترافیک	XM Traffic	ایرماگ	همه به جز : U+066C	بله	بله	بله
۲۰. وحید	XM Vahid	ایرماگ	همه به جز : U+066C	بله	بله	بله
۲۱. یرموک	XM Yermook	ایرماگ	همه به جز : U+066C	بله	بله	بله
۲۲. وستا	XP Vosta	ایرماگ	همه به جز : U+066C	بله	بله	بله
۲۳. زیبا	XP Ziba	ایرماگ	همه به جز : U+066C و U+064A	بله	بله	بله
۲۴. نستعلیق	IranNastaliq	شورای عالی اطلاع رسانی	همه به جز : U+0643 و U+0670 و U+064A و U+2010 و U+FD3E و U+FD3F و U+0649 و U+0656	بله	بله	بله
۲۵. سگو	Segoe UI	مایکروسافت	همه به جز : U+066B و U+066C و U+2010 و U+FD3E و U+FD3F	بله	بله	بله

۲۶	تاهوما	Tahoma	مایکروسافت	همه به جز : U+066C و U+066B و U+FD3F و U+2010 و U+FD3E و U+FD3F و U+0656	بله	بله	بله
۲۷	آریال	Arial (Body CS)	مایکروسافت	همه به جز : U+066C و U+066B و U+FD3F و U+2010 و U+FD3E و U+FD3F و U+0656	بله	بله	بله
۲۸	تیتربی	B Titr	برنا رایانه	نا سازگاری خیلی زیادی دارد	خیر	خیر	خیر
۲۹	شهرزاد	Scheherazade	موسسه غیر انتفاعی sil	به طور کامل	بله	بله	بله
۳۰	لطیف	Lateef	موسسه غیر انتفاعی sil	به طور کامل	بله	بله	بله

جدول ۲: معرفی بعضی کاراکترهای یونی‌کد که در این متن به آنها اشاره شده است

کد	نام رایج فارسی	نام یونیکدی	شکل بر اساس استاندارد یونیکد نسخه ۵.۱
۱	U+0643 حرف «ک» عربی که نباید در متون فارسی استفاده شود.	ARABIC LETTER KAF	ك 0643
۲	U+0649 حرف عربی «ی» بی‌نقطه	ARABIC LETTER ALEF MAKSURA	ى 0649
۳	U+064A حرف «ی» عربی که نباید در متون فارسی استفاده شود.	ARABIC LETTER YEH	ي 064A
۴	U+0656 الف مقصوره پایین فارسی	ARABIC SUBSCRIPT ALEF	ا 0656
۵	U+066B مميز مخصوص عربی (فارسی) که شبیه «اسلش» است نه ممیز و کلمات انگلیسی یا فارسی	ARABIC DECIMAL SEPERATOR	ر 066B
۶	U+066C جدا کننده هزارگان عربی (فارسی) که بر خلاف تصور بالای خط نوشتن می‌نشیند نه پایین آن	ARABIC THOUSANDS SEPERATOR	ر 066C
۷	U+0670 الف مقصوره فارسی	ARABIC LETTER SUPERSCRIPT ALEF	ا 0670
۸	U+2010 این کاراکتر جایگزین دقیق‌تر کاراکتر خط تیره است	HYPHEN	-

۹	U+FD3E	پرافتخار تزیینی چپ	ORNATE LEFT PARENTHESIS	
۱۰	U+FD3F	پرافتخار تزیینی راست	ORNATE RIGHT PARENTHESIS	
۱۱	U+FDEC	علامت مخصوص «ریال»	RIAL SIGN	ریال

نتیجه

تا قبل از اختراع یونی کد، صدها سیستم کدگذاری مختلف وجود داشت. این سیستم‌های کدگذاری، با هم تعارض داشتند؛ یعنی در دو کدگذاری مختلف ممکن بود از اعداد یکسانی برای دو نویسه مختلف، یا از اعداد مختلفی برای نویسه‌های یکسان استفاده شود. هرگاه داده‌ها از کدگذاری‌ها یا محیط‌های مختلف عبور کنند، در معرض خطر تحریف قرار می‌گیرند. یونی کد، استاندارد است که برای کلیه زبان‌های زنده دنیا و خطوط باستانی، بر استانداردهای دیگر برتری دارد؛ به علاوه، با آن می‌توان تمامی زبان‌های دنیا را ترکیب کرد. از این رو متولیان و برنامه‌ریزان صنعت IT کشور به همراه دلسوزان و استادان خط و زبان فارسی باید توجه خود را به این مسئله معطوف کنند که وضعیت فعلی خط فارسی مورد استفاده در نگارش‌های رایانه‌ای، غیراستاندارد و برای آینده متون دیجیتالی خطرآفرین و مشکل‌ساز است و تنها می‌توان با اتکا به کدهای بین‌المللی یونی کد و استانداردهای مصوب و مورد وثوق مؤسسه استاندارد و تحقیقات صنعتی به وضعیت مطلوبی رسید و سپس براساس این کدها، به ایجاد و توسعه فونت‌های زیبا و شکیل فارسی اقدام کرد.

کتابنامه

- مؤسسه استاندارد و تحقیقات صنعتی ایران (<http://www.isiri.rog/std/6219.htm>)
 شورای عالی انفورماتیک کشور (<http://www.shci.ir>)
 سایت ژاپنی وازو (http://www.wazu.jp/gallery/Fonts_Persian2.html)

صفحهٔ دانلود فونت عربی مایکروسافت

(<http://www.microsoft.com/downloads/details.aspx?FamilyID=A83C0E03-8913->)

پروژهٔ فونت فارسی آزاد

(<http://fpf.sourceforge.net/per/index.html>)

<http://www.unicode.org>

<http://unicode.org/standard/translations/persian.html>

هادی صباغ، ماهنامهٔ شبکه، شمارهٔ ۳۹، ص ۱۲۷، سال ۱۳۸۲ (نشر رومیزی در ایران):

<http://www.shabakeh-mag.com/Articles/Show.aspx?n=1000127>

ماهنامهٔ شبکه، بهمن ۱۳۸۳، شمارهٔ ۵۱ (زرنگار و دیگر هیچ):

<http://www.shabakeh-mag.com/Articles/Show.aspx?n=1001849>

علیرضا صالحی، ماهنامهٔ شبکه، خرداد ۱۳۸۳، شمارهٔ ۴۳ (کسب و کاری که دیگر نیست):

<http://www.shabakeh-mag.com/Articles/Show.aspx?n=1000972>