

دانشکده آموزش های الکترونیکی دانشگاه شیراز

درس سیستم های چند رسانه ای

Unicode چیست؟

یونیکد چیست؟

- یونیکد استاندارد است که با استفاده از آن می‌توان برای تمامی زبان‌های زنده و مرده (مثل زبان‌های باستانی) دنیا، یک حرف (کاراکتر) متناظر ایجاد کرد.
- با استفاده از یونیکد یک برنامه‌نویس قادر است برای یک زبان یا یک لهجه خاص، یک فونت ایجاد کند، بدون اینکه برای این کار، حروف (کاراکترهای) زبان یا لهجه مورد نظرش را، در جای مربوط به یک لهجه یا زبان دیگر رونوشت (کپی) کند.
- Anything that you could call a character goes in.

- مزیت این روش آن است که هر "glyph" گلیف " یک کد مخصوص و انحصاری دارد که به غیر خود آن گلیف هیچ گلیف دیگری اجازه استفاده از آن را ندارد .
(به شکل ظاهری حروف گلیف گفته می شود ، این شکل می تواند هندسی - برای زبان های باستانی - یا یک علامت نقاشی شده - مثل زبان هیروگلیف مصر - یا به شکل عدد - یا به شکل هر حرف دیگری باشد)
- این کار باعث جلوگیری از دوبار تکرار شدن یک حرف می شود.
- همچنین باعث می شود ، گلیف (یا حرف یا کاراکتر) مورد نظر به سادگی قابل تشخیص باشد ، مثلا کد 0041 متعلق به حرف A انگلیسی است ، و سایر کاراکترها کدهای مخصوص به خود را دارند .

Unicode vs. Non-Unicode

- **MARC-8) Non-Unicode:**
 - فضاي 8 بيتي براي هر کاراکتر
 - حداکثر 256 کاراکتر
 - 128 تا داراي کد اسكي هستند.
- **UTF-8) Unicode:**
 - افزايش فضا به 16 بيت
 - افزايش تعداد کاراکترهاي قابل کددهي از 256 به 65536!
 - فضا به ميزان کافي براي تعريف گليفيهاي جديد وجود دارد.

بلوکهای یونیکد

- در یونیکد برای اینکه کاراکترهای زبان‌های یک شکل را (مثل فارسی و عربی) دوبار تکرار نکنند، از یک ساختار بلوک بندی شده استفاده کرده‌اند.
- مثل استادیومها که برای دسته‌بندی افراد و مدیریت راحت تر تماشاگران، ساختار بلوک بندی شده در نظر گرفته شده و هر بخش از بخش دیگر متمایز شده است.
- این ساختار از بزرگ شدن یونیکد جلوگیری می‌کنند.
- هر زبان کاراکترهای غیرمشترک را به بلوک کاراکترها اضافه می‌کند.
- در مثال فارسی و عربی کاراکترهای مشابه تکرار نمی‌شوند و از یک کد مشترک هر دو زبان استفاده می‌کنند.
- نکته: در یونیکد هیچ بلوک مشخصی برای زبان فارسی موجود نیست و حروف فارسی در بلوک عربی جای گرفته‌اند.

مثال:

- Basic Latin : 0000..007F;
- Arabic : 0600..06FF;
- Arrows : 2190..21FF;
- Mathematical Operators : 2200..22FF;
- ...

آیا یونیکد فونت است؟

- یونی کد فونت نیست.
- یونی کد یک استاندارد است و فونت‌هایی که پیاده سازی می‌شوند می‌توانند با این استاندارد سازگار باشند و یا نباشند.
- سازگاری یک فونت با یونی کد به آن معنی است که آن فونت خاص ، هنگام پیاده سازی گلیف‌ها، هر گلیف را دقیقاً سر جای خود قرار بدهد .
- *No font provides full Unicode coverage. There's no single font that has all the Unicode characters. Different fonts have different sets of characters, basically depending on which characters the font designer decided to draw.*

- یونی‌کد یک استاندارد است و دارای نسخه های مختلف.

- در هر نسخه از یونی‌کد تعداد کاراکترهای بیشتری (که از قبل در استاندارد، با توجه به زبان‌های مختلف تعریف شده بودند) به محدوده کاراکترهای ایجاد شده می‌پیوندند

- مثلاً در نسخه 6 بعد گلیف‌های زبان‌های باستانی (ایرانی) هم به جمع گلیف‌های ایجاد شده اضافه می‌شوند .

- برای نمونه فونت **Arial** در ویندوز فونتی با حجم بیشتر از 20 مگابایت است که تقریباً بیشتر حروف یونی‌کد که در نسخه پنج وجود دارد را در خود جا داده است .

ساختار داخلي يونيكد

- در يونيكد ، هر گليف براي خود يك كد انحصاري دارد و فقط و تنها فقط همان گليف مجاز به استفاده از آن فضا در بلوك مختص به خود است .
- مثلا عدد 6 فارسي نهي تواند در محل عدد 6 انگليسي قرار بگيرد
- يكي از مهمترين دلایل غيراستاندارد بودن فونتهای فارسي همین است

استقلال از پلتفرم

- یونی‌کد یک استاندارد است و وابسته به هیچ پلتفرم سخت افزاری و یا نرم افزاری نیست.
- یونیکد به هیچ محیط اجرایی وابسته نیست.
- فونت‌های یونی‌کد به ساختار صفحه کلید وابستگی ندارند و مهم نیست که شما از چه آرایشی برای تنظیم کاراکترهای خودتان بروی کیبورد استفاده کرده‌اید .
- هر کاراکتر دارای یک کد انحصاری است و در بلوک مرتبط با خود قرار گرفته و این قضیه هیچ ارتباطی با آرایش صفحه کلید ندارد.

ارتباط با سیستم عامل

- کار کیبورد تولید یک کد به ازای هر کاراکتر است.
- هر وقت روی کلیدی را فشار می دهید، درایور کیبورد (که معمولاً یک فایل **dll** است، مثل **kbdafa.dll**) کد یونی کد معادل آن گلیف خاص را تولید کرده و برای سیستم عامل ارسال می کند.
- سیستم عامل بسته به کد دریافتی اقدام به نمایش و ثبت کلید می کند.

ارتباط با سیستم عامل

- برای اینکه شما بتوانید یونی‌کد را فارسی بنویسید، اول باید این زبان را روی ویندوز یا هر سیستم عامل دیگری فعال کنید.

– با این کاریک **dll** ایجاد می‌شود که ترکیب چیدمان و کدهای یونی‌کد را در خود قرار داده و هنگامی که شما تایپ می‌کنید، بسته به چیدمانی که برای آن تعریف شده، کدهایی را تولید می‌کند.

