



فصل دوم

متن، ابرمتن

متن
کدگذاری
فرمت‌های ذخیره‌سازی
ابرمتن و وب



متن (کدگذاری متن)

ASCII ◀

هفت بیت (۰-۱۲۷) ◀

Unicode ◀

انواع استانداردهای Unicode

UTF-7

UTF-8

UTF-EBCDIC

UTF-16

UTF-32



UTF-8

◀ کد گذاری بیت متغیر

◀ سازگار با اسکی

◀ یک تا چهار بایت را برای یک کاراکتر در نظر می گیرد

◀ One byte ASCII U+00- U+7F

◀ Two bytes Arabic, Hebrew,... U+0080-U+07FF

◀ Three bytes

◀ Four bytes

دامنه	کد UTF-8			
00-7F	0XXXXXXXX			
80-07FF	110XXXXXX	10XXXXXXX		
0800-FFFF	1110XXXXX	10XXXXXXX	10XXXXXXX	
10000-10FFFF	11110XXX	10XXXXXXX	10XXXXXXX	10XXXXXXX

Unicode(Arabic)

دامنه کد ◀

U+0600-U+06FF ◀

اعداد عربی ◀

U+0660-U+0669 ◀

حروف ◀

U+066E-U+06EF ◀

اعداد فارسی - هندی ◀

U+06F0-U+06F9 ◀

اعداد عربی، فارسی

Western (Maghreb, Europe)	Central (Mideast)	Eastern/Indian (Persian, Urdu)
0	۰	۰
1	۱	۱
2	۲	۲
3	۳	۳
4	۴	۴
5	۵	۵
6	۶	۶
7	۷	۷
8	۸	۸
9	۹	۹



صفحه کلید

- ◀ چینش صفحه کلید استاندارد فارسی در ویندوز، با چینش استاندارد ملی ایران، متفاوت است.
- ◀ در حالی که در توزیع‌های گنو/لینوکس این استاندارد (ملی) صفحه کلید به صورت پیش فرض رعایت گردیده است.

حالت عادی

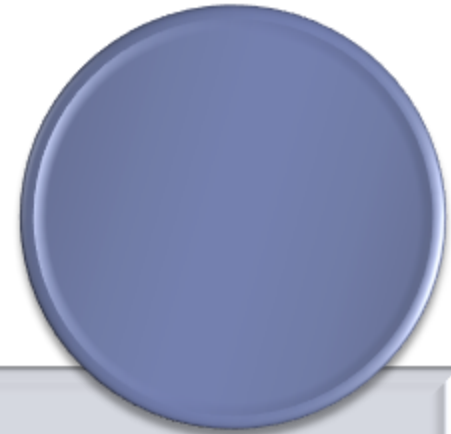
اتصال مجازی	۱	۲	۳	۴	۵	۶	۷	۸	۹	۰	-	=	\	پس بر
جهش	چ	ج	ح	خ	ه	ع	غ	ف	ق	ث	ص	ض		
ورود	گ	ک	م	ن	ت	ا	ل	ب	ی	س	ش			
تبدیل	.	/	و	پ	د	ذ	ر	ز	ط	ظ				
مهار	دگرساز	فاصله	دگرساز راست	مهار										

صفحه کلید استاندارد ملی

UTF-32

◀ ۳۲ بیتی

◀ کدگذاری طول ثابت



متن (فرمت های ذخیره سازی)

- ◀ مخفف document
- ◀ محصول مایکروسافت و مورد استفاده در Microsoft Word
- ◀ فایل های doc معمولا اطلاعاتی بیشتر از متن (مثل Undo) را که در فرمت هایی مثل HTML یا RTF امکان ذخیره سازی ندارد در خود ذخیره می کنند
- ◀ دارای نسخه های مختلف است (۲۰۰۳-۹۷).
- ◀ این فرمت توسط OpenOffice قابل خواندن است. اما ممکن است برخی از اطلاعات سطح بالا مثل سرآیند و ... به دلیل منبع بسته بودن از دست رود.

فرمت docx

- ◀ همراه با Office 2007 به بازار آمد
- ◀ مبتنی بر XML
- ◀ در Office 2007 فرمت جدیدی به نام OpenXML ابداع شد
- ◀ این فرمت از ZIP برای کاهش حجم مستندات استفاده می کند.
- ◀ بنابر اظهارات مایکروسافت با استفاده از این فرمت حجم مستندات Office تا ۷۵٪ کاهش می یابد.

- ◀ **Rich Text Format** مخفف
- ◀ در سال ۱۹۸۷ ابداع شد
- ◀ قابل خواندن توسط انسان (Human-Readable)
- ◀ اغلب پردازشگرها می توانند این فرمت را بخوانند و بنویسند.

◀ کد RTF:

◀ `{\rtf1\ansi{\fonttbl\f0\fswiss Helvetica;}\f0\pard This is some
{\b bold} text.\par }`

◀ به هنگام نمایش اینگونه می شود:

◀ This is some **bold** text.



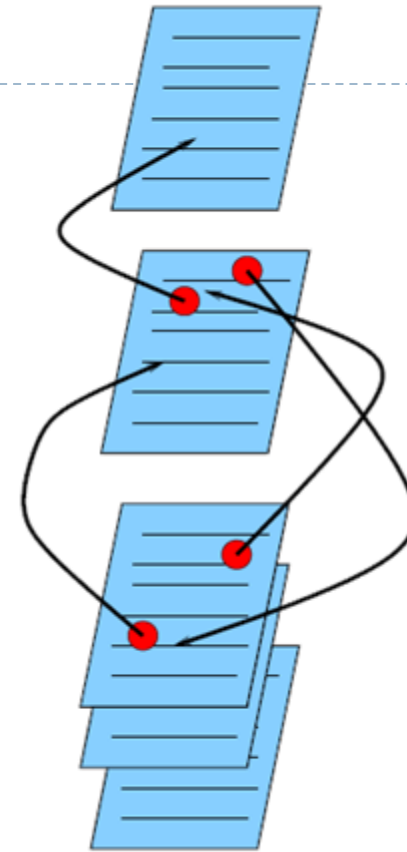
ابر متن و وب

Normal Text



Linear

Hypertext



● "Hot spots"

Nonlinear

- ◀ عملکرد ابرمتن به صورت غیر خطی است.
- ◀ Ted Nelson در سال ۱۹۶۰ ابرمتن را اختراع کرد.

◀ W3C مخفف World Wide Web Consortium است.

این کنسرسیوم کار می کند تا علی رغم اختلاف در فرهنگ، آموزش، توانایی و ... وب در دسترس همه قرار گیرد.

◀ W3C در اکتبر سال ۱۹۹۴ توسط Tim Berners-Lee خالق وب ابداع شد.

◀ برخی از شرکتهای عضو این کنسرسیوم عبارتند از:

IBM, Microsoft, America Online, Apple, Adobe, Sun
Microsystems, ...

◀ W3C بر روی استانداردهای تکنولوژیهای مبتنی بر وب کار می کند.

◀ نام استانداردهای این کنسرسیوم W3C recommendations نام دارد.

W3C recommendations

هدف	تکنولوژی
زبان نشانه‌گذاری جهانی برای وب Hyper Text Markup Language	HTML
ویرایشی مرتب‌تر و سخت گیرانه‌تر نسبت به HTML eXtensible HyperText Markup Language	XHTML
زبانی برای انتقال و ذخیره‌سازی داده eXtensible Markup Language	XML
نحوه نمایش المان‌های HTML را نشان می‌دهد Cascading Style Sheets	CSS
و تکنولوژی‌های مبتنی بر XML نظیر Schema, Xpath, Xquery, SOAP, WSDL, RDF	

زبان توصیف ابرمتن HTML

◀ مخفف:

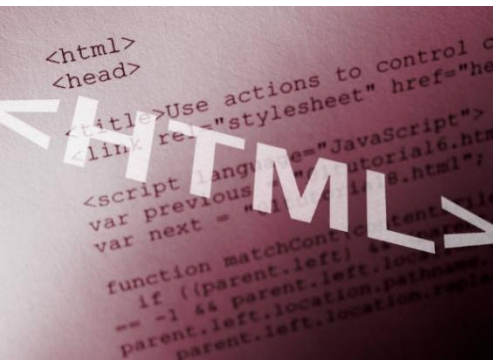
◀ **Hyper Text Markup Language**

◀ یک فایل متنی است که tagهایی در خود دارد

◀ این tagها به مرورگر نحوه نمایش اطلاعات را بازگو می کنند.

◀ HTML از ASCII که قابل انتقال به همه نوع سخت افزار کامپیوتر است استفاده می کند.

◀ نسخه فعلی html نسخه ۴,۰۱ است. W3C روی نسخه ۵ کار می کند.



```
<HTML>
```

```
<HEAD>
```

```
  <TITLE>
```

```
    A sample web page.
```

```
  </TITLE>
```

```
</HEAD>
```

```
<BODY>
```

```
  <P>
```

```
    <B>We can put any text we like here, since this is a  
    paragraph element.</B>
```

```
  </P>
```

```
</BODY>
```

```
</HTML>
```

زبان توصیف توسعه یافته (XML)

Extensible Markup Language ◀ مخفف

- ◀ XML: زبان توصیف و انتقال داده که در آن کاربر می‌تواند tagها را **تعریف** کند.
- ◀ نسخه رایج آن XML 1.0 است که توسط W3C در ۱۹۹۸ استاندارد شده است.
- ◀ مثال از کاربرد XML: بازیابی اطلاعات ذخیره شده از پایگاه داده بر طبق درخواست و پرس و جو (Query) کاربر:

۱. ابتدا از یک نوع مشخصه تعریف مستند (DTD) که از قبل تعریف شده باشد استفاده می‌نماید. (Document Type Definition)
۲. سمت Script server منتظر می‌ماند تا با استفاده از قانون‌های DTD و با توجه به درخواستی که برای استفاده از داده شده است، یک مستند XML تولید کند.
۳. در انتها XML Style Sheet (XSL) با توجه به وسیله‌ای که برای نمایش اطلاعات موجود است برای کاربر ارسال می‌کند.

```
<note>  
<to>Tove</to>  
<from>Jani</from>  
<heading>Reminder</heading>  
<body>Don't forget me this weekend!</body>  
</note>
```



متن (فونت)

فونت‌های کامپیوتری

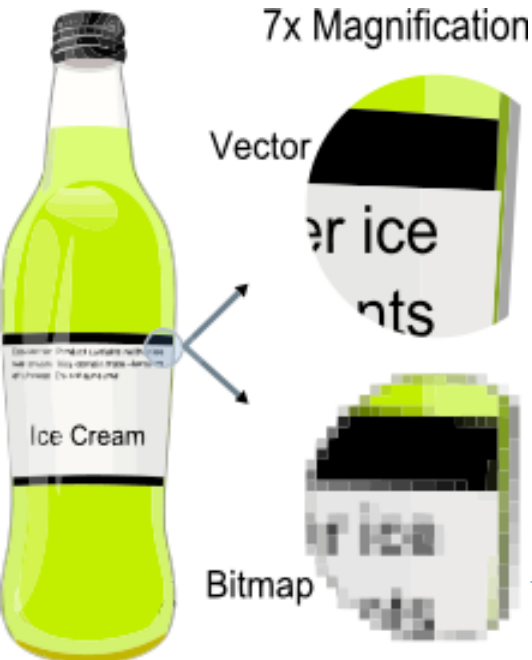
◀ انواع فونت

۱. Bitmap font یا Raster font

هر علامت به صورت آرایه ای از پیکسل‌ها ذخیره شده است. برای هر نمایش یک تصویر به صورت نقش‌بیتی وجود دارد. مثلاً برای هر کاراکتر از فونتی با سه اندازه و ترکیب دلخواهی از سیاه و ایتالیک، ۱۲ تصویر وجود دارد.

این نحوه نمایش سریع و بدون تغییر مقیاس است.

۲. Outline font یا Vector font



Outline font یا Vector font

:TrueType

استانداردی که توسط Apple ابداع شد.
مایکروسافت از آن استفاده می‌کند.

متداول‌ترین نوع فونت

با پسوند ttf

quadratic Bézier curves

:Type1 and Type3

توسط Adobe ابداع شد.

cubic Bézier curves

:FreeType

پیاده‌سازی مستقلی از استاندارد TrueType.

در نسخه‌های لینوکس به کار می‌رود.

- ◀ Web fonts allow Web designers to use fonts that are not installed on the user's computer.
- ◀ When you have found/bought the font you wish to use, just include the font file on your web server, and it will be automatically downloaded to the user when needed.
- ◀ Your "own" fonts are defined within the CSS3 @font-face rule.

```
@font-face {  
  font-family: Yekan;  
  src: url(B Yekan.woff);  
}
```

- ◀ **The Web Open Font Format (WOFF)**
- ◀ WOFF is a font format for use in web pages. It was developed in 2009, and is now a W3C Recommendation. WOFF is essentially OpenType or TrueType with compression and additional metadata. The goal is to support font distribution from a server to a client over a network with bandwidth constraints.
- ◀ **The Web Open Font Format (WOFF 2.0)**
- ◀ TrueType/OpenType font that provides better compression than WOFF 1.0.