

Codificación de caracteres y herramientas en Linux

Recuperación de información
Universidad de Granada 2016/17
José Carlos Entrena Jiménez
Daniel López García
Lothar Soto Palma

Timeline

11:50:00 

Búsqueda: Codificación de caracteres

Tiempo: 5 minutos Web: https://es.wikipedia.org/wiki/Codificaci%C3%B3n_de_caracteres
Conocimiento: Definición del concepto y algunos tipos de codificación .

Tiempo: 10 minutos Web: <https://www.w3.org/International/articles/definitions-characters/index.es>

Conocimiento: Definición Unicode.

Tiempo: 1 minuto Web: http://librosweb.es/libro/xhtml/capitulo_3/codificacion_de_caracteres.html
Conocimiento: Codificación de caracteres en HTML.

12:17:00 

Búsqueda: UTF

Tiempo: 4 minutos Web: <https://es.wikipedia.org/wiki/UTF-8>
Conocimiento: Definición UTF-8

Tiempo: 3 minutos Web: <https://es.wikipedia.org/wiki/UTF-16>
Conocimiento: Definición UTF-16

Timeline

19:50:00 

Búsqueda: editores texto linux formatos texto

Tiempo: 5 minutos Web: <http://proyectopinguino.blogspot.com.es/2009/05/editores-de-texto-para-linux.html>

Conocimiento: Editores de texto de Linux.

Tiempo: 2 minutos Web: <https://jamanajamana.wordpress.com/2012/02/07/tipo-de-editores-de-texto-para-linux/>

Conocimiento: Aparacen los mismos editores.

Tiempo: 10 minutos Web: <https://gsyc.urjc.es/~mortuno/ral/editores.pdf>

Conocimiento: Información más precisa.

20:20:00 

Búsqueda: editores de texto utf8

Tiempo: 2 minutos Web: <http://evamendozaa.blogspot.com.es/2012/02/editores-de-texto.html>

Conocimiento: Aparece la misma información sobre los editores.

Codificación de Caracteres

- La codificación de caracteres es el método que permite convertir un carácter de un lenguaje natural en un símbolo de otro sistema de representación aplicando normas o reglas de codificación.
- Las normas de codificación definen el método para codificar un carácter.
- Ejemplos: Morse, ASCII, Unicode

Unicode

- Unicode es un estándar en el que se definen todos los caracteres necesarios para la escritura de la mayoría de los idiomas hablados en la actualidad que se usan en la computadora.
- Especifica un nombre e identificador numérico único para cada carácter o símbolo.
- Unicode trata los caracteres alfabéticos y símbolos de forma equivalente, lo que significa que se pueden mezclar en un mismo texto sin la introducción de marcas o caracteres de control.
- Unicode define tres formas de codificación bajo el nombre UTF(UTF-8, UTF-16, UTF-32)

0 𐄂 𐄃 𐄄 !
2D30 2D63 2D53 2D4D 21

UTF-16

2D 30 2D 63 2D 53 2D 4D 00 21

UTF-8

0 𐄂 𐄃 𐄄 !
E2 B4 B0 E2 B5 A3 E2 B5 93 E2 B5 8D 21

UTF-32

0 𐄂 𐄃 𐄄 !
00 00 2D 30 00 00 2D 63 00 00 2D 53 00 00 2D 4D 00 00 00 21

Character encodings

Herramientas en Linux

- Los editores de texto crean y modifican ficheros de texto plano.
- Los procesadores de texto crean ficheros de texto con formato e imágenes.



Editores de texto

- Permiten crear y modificar archivos compuestos por texto sin formato.
- Modo texto o modo gráfico.
- La mayoría de editores usan el formato ASCII o UTF-8.
- Ejemplos: vi, gedit, Kwrite, SciTE

Referencias

Codificación de caracteres:

<https://www.w3.org/International/articles/definitions-characters/index.es>

https://es.wikipedia.org/wiki/Codificaci%C3%B3n_de_caracteres

<https://es.wikipedia.org/wiki/UTF-8>

Herramientas Linux:

<https://gsyc.urjc.es/~mortuno/ral/editores.pdf>

<http://proyectopinguino.blogspot.com.es/2009/05/editores-de-texto-para-linux.html>