



Nombre: .....

Fecha:     /09/2010

Grupo: 1 ☐ 2 ☐ 3 ☐ 4 ☐

## PRÁCTICA 1

## INFORMÁTICA BÁSICA, CODIFICACIONES

Entre los objetivos de la práctica se encuentra el trabajar con distintas normas de codificación disponibles en nuestro sistema. Veremos el resultado de guardar ciertas informaciones en nuestro sistema con respecto a alguna codificación, qué sucede cuando intentamos recuperarlas con codificaciones distintas, formas de cambiar las páginas de códigos o las posibilidades y consecuencias de cambiar la configuración regional del sistema.

1. Manejaremos en esta práctica las siguientes direcciones de Internet.

¿Qué es ASCII? Tabla ASCII y manejo de binario, hexadecimal y decimal:

[http://es.wikipedia.org/wiki/American\\_Standard\\_Code\\_for\\_Information\\_Interchange](http://es.wikipedia.org/wiki/American_Standard_Code_for_Information_Interchange)

Tablas de códigos:

<http://www.microsoft.com/globaldev/reference/cphome.mspx>

¿Qué es Unicode?

<http://es.wikipedia.org/wiki/Unicode>

<http://msdn.microsoft.com/es-es/goglobal/bb688113.aspx>

¿Qué es UTF-8?

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

Consorcio Unicode:

<http://www.unicode.org/>

Lista de caracteres Unicode (de los 65536 primeros caracteres):

[http://en.wikibooks.org/wiki/Unicode/Character\\_reference/0000-FFFF](http://en.wikibooks.org/wiki/Unicode/Character_reference/0000-FFFF)

Representación de algunos de los caracteres Unicode en codificación UTF-8 en hexadecimal:

<http://www.sslug.dk/~chlor/utf-8/>

2. Comentamos las direcciones anteriores y los conceptos que queremos trabajar.

3. Preparamos una carpeta con el nombre "Practica" en "Mis Documentos", donde guardaremos los ficheros de esta práctica.

4. Ejecutamos el bloc de notas ("notepad") y escribimos en él el siguiente texto:

```
Esta fila es sólo de caracteres ASCII?  
¿Esta línea tiene algún carácter no ASCII?  
Ñandú, garçon, cigüeña
```

5. Guardamos el fichero anterior ("Guardar como") con tres codificaciones distintas, ANSI, Unicode (UTF-16) y UTF-8 con los nombres "textoAnsi", "textoUnicode" y "textoUtf8", respectivamente (observamos los "tipos de formato" en que sabe codificar información "notepad" y lo comentamos).

6. Ejecutamos el programa "wordpad". Escribimos el mismo texto anterior. Vamos a "Archivo: Guardar como" y observamos los "tipos de formato" en

que sabe escribir. Lo comentamos. Guardamos el archivo como "Documento de texto formato MSDOS" con el nombre "textoDos".

7. Vamos a la carpeta y, pulsando sobre cada uno de los archivos con el botón derecho y seleccionando "Propiedades", observamos cuánto ocupan. ¿Qué ficheros ocupan menos? ¿Cuál es la diferencia entre las distintas codificaciones? ¿Qué fichero ocupa más? ¿Cuál de las codificaciones usadas es de longitud variable?

8. Ejecutamos el programa "Pspad". Vamos a "Archivo: Abrir en hexadecimal" y abrimos cada uno de los 4 ficheros que hemos guardado anteriormente. Observa las tablas de codificaciones presentadas antes. Trata de responder a las siguientes preguntas. ¿Qué longitud de palabra en bits tiene cada una de las codificaciones MSDOS, ANSI, UTF-8 y Unicode? ¿Qué símbolos del texto tienen la misma representación en las cuatro codificaciones? ¿Qué símbolos no? Sobre los símbolos que no comparten representación, ¿tienen la misma representación en Unicode y en UTF-8? ¿Qué sucedería si guardas un fichero con la codificación UTF-8 y lo abres (o lo visualizas) con la codificación UTF-16 (o Unicode)? ¿Se vería correctamente?

9. Ejecuta un navegador (Explorer, Firefox...). Abre una página al azar. ¿En qué codificación se habrán guardado los caracteres de dicha página? Intenta observar el código fuente de la misma y comprobarlo. Ahora abre el menú "Página: Codificación" en Explorer o "Ver: Codificación de caracteres" en Firefox. Observa todas las codificaciones que contiene. ¿Qué pasa si cambiamos la codificación con respecto a la cual estamos leyendo la página? ¿Qué caracteres cambian? ¿Cuáles se siguen mostrando bien?

10. Trabajaremos ahora con la introducción de caracteres que no están en el teclado. Esto lo podemos hacer a través del "Mapa de caracteres", que encontrarás en "Accesorios – Herramientas del Sistema". Observa las distintas tablas de codificaciones de que dispone tu sistema, y sus caracteres correspondientes. ¿De cuántos caracteres dispone "DOS - Europa Occidental" (tabla 850)? ¿Y "Windows - Occidental" (tabla 1252)? ¿Y Unicode? Relaciona la respuesta con las longitudes de palabra correspondiente (recuerda que un "byte" permite almacenar  $2^8 = 256$  valores distintos).

Selecciona la tabla de códigos "DOS – Europa Occidental" y alguno de los caracteres especiales de la misma (entre el 129 y el 256). Cópialo en tu fichero "textoDOS.txt". Ejecuta una consola de comandos de MSDOS (puedes escribir "cmd" en la ventana "Buscar programas y archivos"). Muestra el fichero "textoDOS.txt" por medio del comando "type". ¿Se muestran correctamente los caracteres? Existe otra forma de insertar caracteres especiales (en realidad, cualquier carácter). La puedes encontrar en el Anexo de la práctica. Compruébala en la propia consola de MSDOS y en el fichero "textoDOS.txt" (con "wordpad") incluyendo algunos caracteres que no estén en el teclado.

11. Desde la propia consola de MSDOS podemos modificar la página de caracteres que se está usando en la consola. Para ello debes utilizar el

comando "chcp". Algunas de las tablas estándar que suelen estar instaladas son 850 (tabla por defecto en la consola para nuestra región, "DOS – Europa Occidental"), 437 ("DOS – USA"), 1252. Comprueba su disponibilidad y la capacidad para introducir en la consola caracteres especiales por medio de la combinación "Alt + codificación decimal" (puedes encontrar una lista más detallada de tablas y los caracteres que ocupan cada posición en <http://msdn.microsoft.com/en-us/library/cc195051.aspx> y más detalles sobre la introducción de caracteres por medio de su codificación decimal en el Anexo al final de la práctica).

12. Comentamos las ideas sobre el papel del S.O. en este caso y comentamos qué puede suceder en otros sistemas operativos. Si está Linux instalado en tu ordenador, trata de comprobar qué tabla de código usa por defecto.

13. Ve a la dirección

<http://www.aq.upm.es/Departamentos/Fisica/agmartin/webpublico/latex/FAQ-CervanTeX-main.html> y entra en cada uno de los enlaces de abajo. ¿Qué sucede al leerlos con el navegador? ¿Por qué? ¿Qué sucederá en otro navegador o sobre otro S.O.?

14. Ahora nos bajamos a nuestro ordenador las versiones en latin1, cp850 y cp437. Observamos que, con toda probabilidad, la latin1 se referirá a ISO-8859-1 (Latin 1) y no a Windows-1252 (Latin 1). Lo comprobamos y observamos las posibles diferencias.

#### Anexo – Ayuda

Introducción de caracteres por medio de su posición en la tabla de códigos:

1. Mantener presionada la tecla ALT y, después, presionar las teclas del teclado numérico que representan el valor del código decimal del carácter que se desea introducir.
2. Una vez termine de escribir, suelte la tecla ALT.
3. "notepad" o la consola de MSDOS generan el carácter especificado.

#### Notas

Si el primer dígito que se escribe es 0, el valor se reconoce como un punto de código, o valor de carácter, del idioma actual (el "idioma actual" lo puedes seleccionar desde la barra de herramientas). Por ejemplo, cuando el idioma actual es el inglés estadounidense (página de códigos 1252: Latin 1 de Windows), al presionar la tecla ALT y escribir después 0163 en el teclado numérico, se produce £, el símbolo de la libra (U+00A3). Cuando el idioma actual es el ruso (página de códigos 1251: cirílico de Windows), la misma secuencia de teclas produce la letra mayúscula cirílica JE (U+0408).

Si el primer dígito que escribe es un número del 1 al 9, el valor se reconoce como un punto de código en la página de códigos OEM del sistema. El resultado difiere en función del idioma del sistema Windows especificado en Configuración regional y de idioma del Panel de control. Por ejemplo, si el idioma del sistema es inglés estadounidense, la página de códigos es 437 (MS-DOS Latín EE.UU.) y si se presiona la tecla ALT y se escribe después 163 en el teclado numérico, se produce ú (U+00FA, letra minúscula latina U con acento). Si el idioma del sistema es griego (página de códigos OEM 737 MSDOS griego), la misma secuencia produce la letra minúscula griega MU (U+03BC).

Nota: "pspad" es software libre disponible bajo licencia GNU. Puedes descargarlo desde la dirección <http://www.pspad.com/>