



CFGS DAM-DAW

ACTIVITAT

Codificación de caracteres

M01 – SISTEMES INFORMÀTICS

UF1 – Instal·lació, configuració i explotació del sistema

NF1 – Introducció als sistemes d'informació

NOM ALUMNE: Eina Coma Bages

Objectius generals de l'activitat

Trabajar con algunas de las distintas normas de codificación disponibles en nuestro sistema. Veremos el resultado de guardar ciertas informaciones en nuestro sistema con diferentes normas de codificación y qué sucede cuando intentamos recuperarlas.

Desenvolupament de la l'activitat

En esta práctica vamos a utilizar información de algunas de las siguientes direcciones de Internet.

Tablas de códigos:

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

¿Qué es Unicode?

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

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

¿Qué es UTF-8?

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

Consortio Unicode:

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

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

http://en.wikibooks.org/wiki/Unicode/Character_reference/0000-0FFF

Calculadora de representación de los caracteres Unicode en codificación UTF-8:

<http://www.ltg.ed.ac.uk/~richard/utf-8.cgi>

1. A continuación se da la posición de 6 caracteres correspondientes a la tabla ASCII reducida. Con la ayuda de las teclas “ALT + Teclado Numérico” haya los caracteres a los que se corresponden. Rellena el resto de columnas de la tabla.

Posición en la taula ASCII	Carácter ASCII	Codificación ASCII reducido (7 bits)							Codificación ASCII reducido (blanco) + bit de paridad par (gris) (8 bits en total)							
65	A	1	0	0	0	0	0	1	0	1	0	0	0	0	0	1
97	a	1	1	0	0	0	0	1	1	1	0	0	0	0	0	1
35	#	0	1	0	0	0	1	1	1	0	1	0	0	0	1	1
57	/	0	1	1	1	0	0	1	0	0	1	1	1	0	0	1
64	@	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0
126	~	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0

2. A partir de la tabla siguiente, y de la información contenida en la primera columna (Codificación binaria de caracteres ASCII) calcula la segunda columna (Posición hexadecimal) y tercera columna (Posición decimal). Sabiendo la posición decimal que ocupa cada carácter, y con la ayuda de las teclas “Alt + Teclado Numérico”, haya el carácter representado y el mensaje codificado. Localiza las faltas de ortografía y propón su corrección.

¿La codificación de caracteres ASCII utiliza el bit de paridad par para el control de errores?

Comprueba si la codificación ASCII de los caracteres numéricos que aparecen en el mensaje coincide con el valor obtenido de la conversión de decimal a binario.

Codificación ASCII	Posición hexadecimal	Posición decimal	Carácter representado (Mensaje)	La codificación ASCII habría de ser (falta ortográfica):
0100 1000	48	72	H	
0110 0001	61	97	a	
0111 0011	73	115	s	
0010 0000	20	32		
0111 0000	70	112	p	
0110 0001	61	97	a	
0110 0111	67	103	g	
0110 0001	61	97	a	
0110 0100	64	100	d	
0011 0000	30	48	0	0110 1111
0010 0000	20	32		
0011 0011	33	51	3	
0011 0110	36	54	6	
0011 0000	30	48	0	
0010 0000	20	32		
0110 0100	64	100	d	
0110 1111	6F	111	o	
0010 0000	20	32		
0111 0100	74	116	t	
0110 0001	61	97	a	
0111 1000	71	113	s	
0110 0001	61	97	a	
0111 0011	73	115	s	
0010 0001	21	33	!	

3. ¿Cuál es el equivalente en Ubuntu a la ayuda de las teclas “Alt + Teclado Numérico” de Windows, que permite escribir un carácter conocido su posición decimal en la tabla de caracteres ASCII? Obtén el carácter “:” sabiendo que valor hexadecimal que te indica la tabla siguiente:

$ \begin{array}{c} b_7 \xrightarrow{\hspace{1cm}} \\ b_6 \xrightarrow{\hspace{1cm}} \\ b_5 \xrightarrow{\hspace{1cm}} \\ \text{Bits} \quad b_4 \downarrow \quad b_3 \downarrow \quad b_2 \downarrow \quad b_1 \downarrow \\ \begin{array}{c c} \text{Column} \rightarrow & \\ \text{Row} \downarrow & \end{array} \end{array} $					0	0	0	0	1	0	1	1	0	1	1	1	1
					0	0	0	1	0	1	1	0	0	1	1	0	1
					0	1	2	3	4	5	6	7					
0					0	NUL	DLE	SP	0	@	P	`	p				
0					1	SOH	DC1	!	1	A	Q	a	q				
0					2	STX	DC2	"	2	B	R	b	r				
0					3	ETX	DC3	#	3	C	S	c	s				
0					4	EOT	DC4	\$	4	D	T	d	t				
0					5	ENQ	NAK	%	5	E	U	e	u				
0					6	ACK	SYN	&	6	F	V	f	v				
0					7	BEL	ETB	'	7	G	W	g	w				
1					8	BS	CAN	(8	H	X	h	x				
1					9	HT	EM)	9	I	Y	i	y				
1					10	LF	SUB	*	:	J	Z	j	z				
1					11	VT	ESC	+	;	K	[k	{				
1					12	FF	FC	,	<	L	\	l					
1					13	CR	GS	-	=	M]	m	}				
1					14	SO	RS	.	>	N	^	n	~				
1					15	SI	US	/	?	O	_	o	DEL				

Sin dejar de presionar **"CTRL+SHIFT+u"**, presiona en el teclado el valor hexadecimal de la letra o símbolo en el código ASCII. Así, por ejemplo, para obtener el carácter ":", sin dejar de presionar **"CTRL+SHIFT+u"**, presiona en el teclado **"3a"**, que es el valor hexadecimal de la letra o símbolo ":" en el código ASCII.

4. Rellena la tabla siguiente, sabiendo que todos los caracteres representados son códigos UTF-8 de UNICODE.

Carácter	Punto de código	Código Unicode binario	Nº bytes utilizados	Carácter codificado en UTF-8 (hexadecimal)
a	U+0061	0110 0001	1	01100001
ç	U+00E7	1110 0100	2	11000011 10100100
ñ	U+00F1	1111 0001	2	11000011 10110001
🐼	U+1F4A9	0001 1111 0100 1010 1001	4	11110000 10011111 10010010 10101001
Ali Gali Baluda	U+1885	0001 1000 1000 0101	3	11100001 10100010 10000101
₪	U+1CDD0 ¹	0001 1100 1101 1101 0000	4	11110000 10011100 10110111 10010000
₪	U+16080 ²	0001 0110 0000 1000 0000	4	11110000 10010110 10000010 10000000

1 (a la proposta L2/20-181)

2 Té l'espai assignat al SMP, però encara no forma part de l'estàndard.

(1) Escribe un carácter cualquiera del **idioma klingon**, lengua construida como lengua vernácula de la raza Klingon en el universo de Star Trek. Fue diseñado con un orden de palabras tipo Objeto Verbo Sujeto (OVS) para hacerlo menos intuitivo y darle un aspecto más alienígena.

(2) Escribe un carácter cualquiera del **idioma Élfico** (del Señor de los anillos)

PARTE OPTATIVA DE LA PRÁCTICA

5. Crea una carpeta con el nombre “Codificación de caracteres”, donde guardaremos los ficheros de esta práctica. Ejecuta el bloc de notas (“notepad”) y escribe en él la letra “a” minúscula. Guarda el fichero anterior (“Guardar como”) con las cuatro codificaciones posibles: ANSI, Unicode (UTF-16), Unicode Big Endian (UTF-16BE) y UTF-8 con los nombres “a_Ansi”, “a_UTF16”, “a_UTF16BE” y “a_UTF8”, respectivamente (observa los “tipos de formato” en se puede codificar la información “notepad”).

Llamamos “longitud de palabra” o “longitud de carácter” a la cantidad de memoria que ocupa cada carácter propio de una codificación cuando el mismo es almacenado en memoria.

Entra en la carpeta “Codificación de caracteres” y, pulsando sobre cada uno de los archivos con el botón derecho y seleccionando “Propiedades”, observa cuántos bytes ocupan. Rellena la tabla siguiente:

Formato de la codificación	Tamaño que HD	Tamaño que ocupa el archivo	Codificación hexadecimal ASCII
ANSI			
Unicode (UTF-16)			
Unicode Big Endian (UTF-16BE)			
UTF-8			

¿Por qué no coincide el tamaño de los ficheros con el tamaño que necesitan para ser almacenados en el disco duro?

6. Instala en tu ordenador la aplicación “**Pspad**”, uno de los muchos editores hexadecimales que hay. Puedes descárgalo para su instalación en el ordenador desde la página <http://www.pspad.com/es/download.php>.

Ejecuta Pspad. Vamos a “Archivo: Abrir en hexadecimal” y abrimos cada uno de los 4 ficheros que hemos guardado anteriormente. Rellena la siguiente tabla teniendo en cuenta que dos símbolos hexadecimales pueden corresponder a un único carácter (un mismo byte).

Formato de la codificación del archivo	Codificación hexadecimal del archivo	Código little-endian	Código big-endian	Código BOM	Código del carácter a
a_ANSI					
a_Unicode (UTF-16)					
a_Unicode Big Endian (UTF-16BE)					
a_UTF-8					

7. ¿Puedes representar en el bloc de notas ("notepad") algún carácter de que no esté incluido en la tabla ASCII estándar? En caso afirmativo, realiza una tabla análoga a la de la pregunta anterior, pero en este caso para el carácter "€".
8. Con la ayuda de la tabla de codificación ASCII escribe el código esperado para la palabra "hola". Escribe la palabra hola en el bloc de notas ("notepad") y guárdalo en los todos los formatos que te ofrece la aplicación. Comprueba la diferencia entre el código esperado para la la palabra "hola" y el que te ofrece el editor hexadecimal. Explica la diferencia entre la codificación big y little endian.
9. Abre en diferentes navegadores (Explorer, Firefox, Chrome...) la página web del instituto. ¿En qué codificación se han guardado los caracteres de dicha página? Intenta observar el código fuente de la misma y comprobarlo.
Abre el menú "Página: Codificación" en los navegadores Explorer, Firefox y Chrome. Observa todas las codificaciones que contienen. ¿Qué sucede si cambiamos la codificación por defecto con la estamos visualizando la página? ¿Qué caracteres cambian? ¿Cuáles se siguen mostrando correctamente? Entre las distintas versiones, compara la visualización actual de la página web con respecto a la codificación UTF-16 y explica el resultado obtenido.