

Sistemas de información y base de datos

CLASE 4

UNTREF
2016

Emilio Platzer
emilioplatzer@gmail.com

Licencias

Trabajemos con casos de prueba

<https://github.com/untref-SlyBD-2016/alumnos/.../clase4.md>

Licencias

Open source (código abierto)

- centrado en aspectos prácticos
- permite el uso comercial

puntos en común

- el código fuente puede ser leído por cualquiera
- está permitido modificarse
- debe reconocer al autor

Free software (software libre)

- centrado en aspectos éticos y filosóficos
- obliga a liberar a los productos derivados

{?}

choosealicense.com



Permissions

- Commercial Use
- Distribution
- Modification
- Private Use

Conditions

- License and Copyright Notice

Limitations

- Hold Liabile

- Patent Use

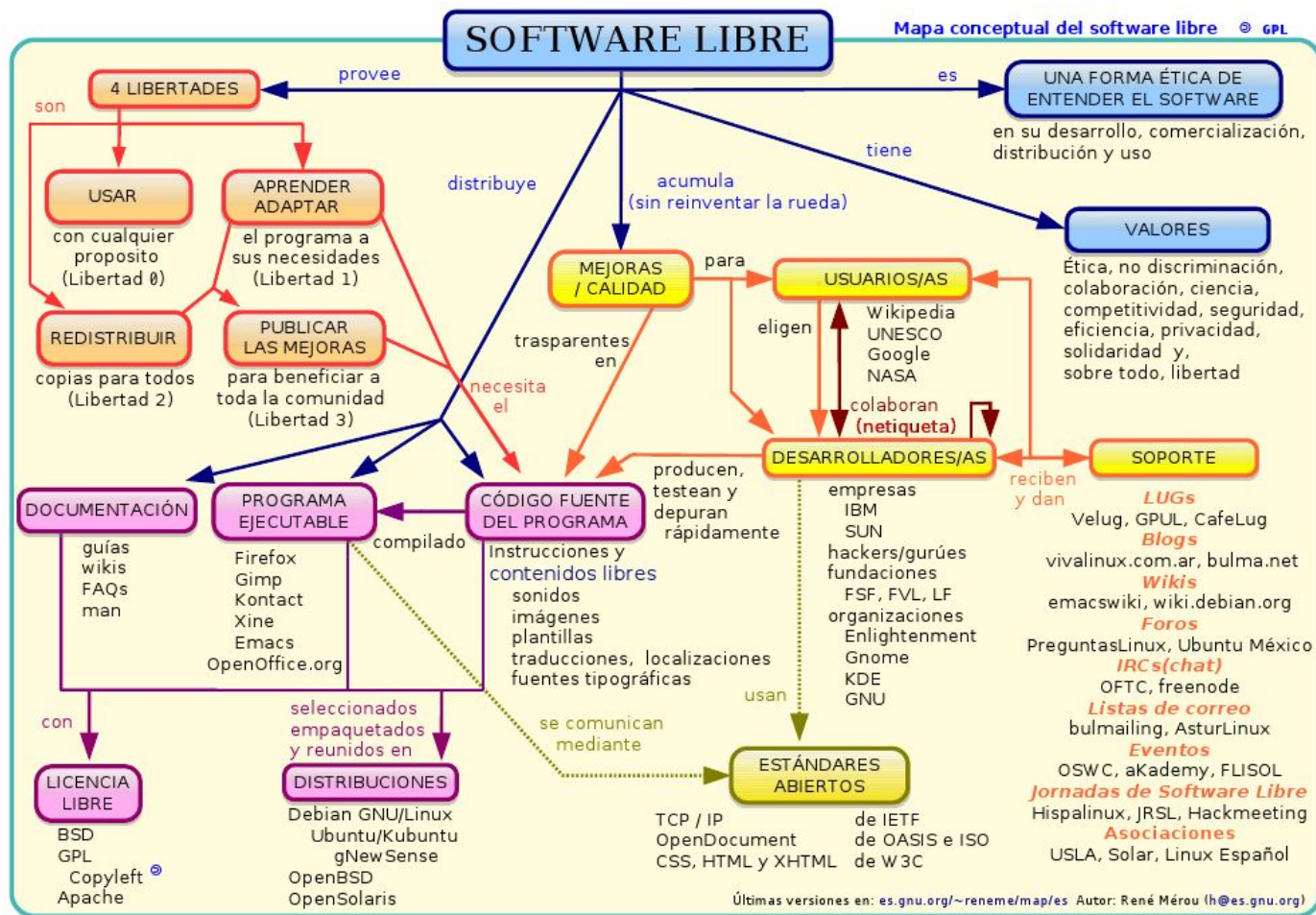
- State Changes

características
agregadas a la
licencia anterior

- Disclose Source

- Same License

Últimas versiones en: es.gnu.org/~reneme/map/es Autor: René Mérou (h@es.gnu.org)



íconos para las licencias



Dominio público



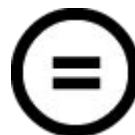
Libre



Creative commons



Copyleft



Copyright

formatos de archivos: codificación de caracteres

ASCII7

	-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-A	-B	-C	-D	-E	-F
0-		0001	0002	0003	0004	0005	0006	0007	0008	0009	000A	000B	000C	000D	000E	000F
1-	0010	0011	0012	0013	0014	0015	0016	0017	0018	0019	001A	001B	001C	001D	001E	001F
2-		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3-	0020	0021	0022	0023	0024	0025	0026	0027	0028	0029	002A	002B	002C	002D	002E	002F
4-	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
5-	0030	0031	0032	0033	0034	0035	0036	0037	0038	0039	003A	003B	003C	003D	003E	003F
6-	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
7-	0040	0041	0042	0043	0044	0045	0046	0047	0048	0049	004A	004B	004C	004D	004E	004F
8-	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
9-	0050	0051	0052	0053	0054	0055	0056	0057	0058	0059	005A	005B	005C	005D	005E	005F
10-	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
11-	0060	0061	0062	0063	0064	0065	0066	0067	0068	0069	006A	006B	006C	006D	006E	006F
12-	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
13-	0070	0071	0072	0073	0074	0075	0076	0077	0078	0079	007A	007B	007C	007D	007E	007F

formatos de archivos: codificación de caracteres

UNICODE

	008	009	00A	00B	00C	00D	00E	00F		010	011	012	013	014	015	016	017
0	XXX 0080	DCS 0090	NB SP 00A0	° 00B0	À 00C0	Đ 00D0	à 00E0	đ 00F0		Ā 0100	Ď 0110	Ġ 0120	Ĭ 0130	Ĳ 0140	Ŏ 0150	Š 0160	Ů 0170
1	XXX 0081	PU1 0091	ı 00A1	± 00B1	Á 00C1	Ñ 00D1	á 00E1	ñ 00F1		ā 0101	ď 0111	ġ 0121	ĳ 0131	Ľ 0141	ő 0151	š 0161	ů 0171
2	BPH 0082	PU2 0092	¢ 00A2	² 00B2	Â 00C2	Ò 00D2	â 00E2	ò 00F2		Ǻ 0102	Ě 0112	Ģ 0122	Ĳ 0132	Ł 0142	œ 0152	Ť 0162	Ů 0172
3	NBH 0083	STS 0093	£ 00A3	³ 00B3	Ã 00C3	Ó 00D3	ã 00E3	ó 00F3		ǻ 0103	ě 0113	ģ 0123	ĳ 0133	Ł 0143	æ 0153	ţ 0163	ı 0173
4	IND 0084	CCH 0094	¤ 00A4	´ 00B4	Ä 00C4	Ô 00D4	ä 00E4	ô 00F4		Ą 0104	Ě 0114	Ĥ 0124	Ĵ 0134	ń 0144	Ŕ 0154	Ť 0164	Ű 0174
5	NEL 0085	MW 0095	¥ 00A5	µ 00B5	Å 00C5	Õ 00D5	å 00E5	õ 00F5		ą 0105	ě 0115	ĥ 0125	ĵ 0135	Ń 0145	ŕ 0155	ť 0165	ŵ 0175
6	SSA 0086	SPA 0096	ı 00A6	¶ 00B6	Æ 00C6	Ö 00D6	æ 00E6	ö 00F6		Ć 0106	Ě 0116	Ħ 0126	Ɔ 0136	Ŗ 0146	ŕ 0156	Ŧ 0166	Ŷ 0176
7	ESA 0087	EPA 0097	§ 00A7	· 00B7	Ç 00C7	× 00D7	ç 00E7	÷ 00F7		ć 0107	ě 0117	ħ 0127	ķ 0137	Ň 0147	ŕ 0157	ţ 0167	ŷ 0177
8	HTS 0088	SOS 0098	¨ 00A8	ˆ 00B8	È 00C8	Ø 00D8	è 00E8	ø 00F8		Ĉ 0108	Ė 0118	Ĩ 0128	κ 0138	ň 0148	Ř 0158	Ũ 0168	ÿ 0178

1242
latin1

852
latin2

ANSI

	-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-A	-B	-C	-D	-E	-F
8-	€ 20AC	‚ 0081	‚ 201A	ƒ 0192	” 201E	… 2026	† 2020	‡ 2021	^ 02C8	% 2030	Š 0160	< 2039	Œ 0152		Ž 017D	
9-	‘ 0090	’ 2018	‚ 2019	“ 201C	” 201D	• 2022	— 2013	™ 2122	~ 02DC		> 0161	œ 0153		ž 0090	ÿ 017E	ÿ 017F
A-	‚ 00A0	‚ 00A1	‚ 00A2	‚ 00A3	‚ 00A4	‚ 00A5	‚ 00A6	‚ 00A7	‚ 00A8	‚ 00A9	‚ 00AA	‚ 00AB	‚ 00AC	‚ 00AD	‚ 00AE	‚ 00AF
B-	‚ 00B0	‚ 00B1	‚ 00B2	‚ 00B3	‚ 00B4	‚ 00B5	‚ 00B6	‚ 00B7	‚ 00B8	‚ 00B9	‚ 00BA	‚ 00BB	‚ 00BC	‚ 00BD	‚ 00BE	‚ 00BF
C-	À 00C0	Á 00C1	Â 00C2	Ã 00C3	Ä 00C4	Å 00C5	Æ 00C6	Ç 00C7	È 00C8	É 00C9	Ê 00CA	Ë 00CB	Ì 00CC	Í 00CD	Î 00CE	Ï 00CF
D-	Ð 00D0	Ñ 00D1	Ò 00D2	Ó 00D3	Ô 00D4	Õ 00D5	Ö 00D6	× 00D7	Ø 00D8	Ù 00D9	Ú 00DA	Û 00DB	Ü 00DC	Ý 00DD	Þ 00DE	ß 00DF
E-	à 00E0	á 00E1	â 00E2	ã 00E3	ä 00E4	å 00E5	æ 00E6	ç 00E7	è 00E8	é 00E9	ê 00EA	ë 00EB	ì 00EC	í 00ED	î 00EE	ï 00EF
F-	ð 00F0	ñ 00F1	ò 00F2	ó 00F3	ô 00F4	õ 00F5	÷ 00F6	ø 00F7	ù 00F8	ú 00F9	û 00FA	ü 00FB	ý 00FC	ÿ 00FD	þ 00FE	ÿ 00FF
8-	Ç 00C7	ü 00FC	é 00E9	â 00E2	ä 00E4	û 00FB	ç 00E7	ı 0107	ë 00EB	ı 0142	ő 0151	ı 00EE	ı 0179	ı 00C4	ı 00C6	ı 00C6
9-	É 00C9	ı 0139	ı 013A	ı 00F4	ı 00F6	ı 013D	ı 013E	ı 015A	ı 015B	ı 00D6	ı 00DC	ı 0164	ı 0165	ı 0141	ı 00D7	ı 01D0
A-	á 00E1	ı 00ED	ı 00F3	ı 00FA	ı 0104	ı 0105	ı 017D	ı 017E	ı 0118	ı 0119	ı 00AC	ı 017A	ı 01C0	ı 015F	ı 00AB	ı 00EB
B-	ı 2591	ı 2592	ı 2593	ı 2594	ı 2595	ı 2596	ı 2597	ı 2598	ı 2599	ı 259A	ı 259B	ı 259C	ı 259D	ı 259E	ı 259F	ı 259F
C-	ı 2514	ı 2515	ı 2516	ı 2517	ı 2518	ı 2519	ı 251A	ı 251B	ı 251C	ı 251D	ı 251E	ı 251F	ı 251F	ı 251F	ı 251F	ı 251F
D-	ı 0111	ı 0112	ı 0113	ı 0114	ı 0115	ı 0116	ı 0117	ı 0118	ı 0119	ı 011A	ı 011B	ı 011C	ı 011D	ı 011E	ı 011F	ı 011F
E-	ı 00D3	ı 00D4	ı 00D5	ı 00D6	ı 00D7	ı 00D8	ı 00D9	ı 00DA	ı 00DB	ı 00DC	ı 00DD	ı 00DE	ı 00DF	ı 00E0	ı 00E1	ı 00E2
F-	ı 00A0	ı 00A1	ı 00A2	ı 00A3	ı 00A4	ı 00A5	ı 00A6	ı 00A7	ı 00A8	ı 00A9	ı 00AA	ı 00AB	ı 00AC	ı 00AD	ı 00AE	ı 00AF

8-	€ 20AC		,	<i>f</i>	”	...	†	‡	^	‰	Š	<	Œ	
	0081	201A	0192	201E	2026	2020	2021	02C6	2030	0160	2039	0152	008D	
9-		‘	’	“	”	●	—	—	~	™	š	>	œ	
	0090	2018	2019	201C	201D	2022	2013	2014	02DC	2122	0161	203A	0153	009D
A-		¡	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	-
	00A0	00A1	00A2	00A3	00A4	00A5	00A6	00A7	00A8	00A9	00AA	00AB	00AC	00AD
B-	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½
	00B0	00B1	00B2	00B3	00B4	00B5	00B6	00B7	00B8	00B9	00BA	00BB	00BC	00BD
C-	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í
	00C0	00C1	00C2	00C3	00C4	00C5	00C6	00C7	00C8	00C9	00CA	00CB	00CC	00CD
D-	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý
	00D0	00D1	00D2	00D3	00D4	00D5	00D6	00D7	00D8	00D9	00DA	00DB	00DC	00DD
E-	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í
	00E0	00E1	00E2	00E3	00E4	00E5	00E6	00E7	00E8	00E9	00EA	00EB	00EC	00ED
F-	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý
	00F0	00F1	00F2	00F3	00F4	00F5	00F6	00F7	00F8	00F9	00FA	00FB	00FC	00FD

Ç 00C7	ü 00FC	é 00E9	â 00E2	ä 00E4	û 016F	ć 0107	ç 00E7	ł 0142	ë 00EB	Œ 0150	œ 0151	î 00EE	Ž 0179	Ä 00C4	Ć 0106
É 00C9	Ł 0139	Í 013A	ô 00F4	ö 00F6	Ľ 013D	ĭ 013E	Ś 015A	ś 015B	Ö 00D6	Ü 00DC	Ť 0164	ť 0165	Ł 0141	×	č 010D
á 00E1	í 00ED	ó 00F3	ú 00FA	Ą 0104	ą 0105	Ž 017D	ž 017E	Ę 0118	ę 0119	€ 20AC	ž 017A	Č 010C	ş 015F	«	» 00BB
 2591	 2592	 2593	 2502	└ 2524	Á 00C1	Â 00C2	Ě 011A	Ş 015E	≡ 2563	 2551	≡ 2557	≡ 255D	Ž 017B	ž 017C	└ 2510
L 2514	┐ 2534	└ 252C	└ 251C	— 2500	┐ 253C	Ǻ 0102	ǻ 0103	┐ 255A	┐ 2554	≡ 2569	≡ 2566	┐ 2560	= 2550	≡ 256C	⌘ 00A4
đ 0111	Đ 0110	Ǻ 010E	ǻ 00CB	ǻ 010F	Ǻ 0147	Í 00CD	Î 00CE	ě 011B	└ 2518	└ 250C	■ 2588	■ 2584	Ť 0162	Ů 016E	■ 2580
Ó 00D3	ß 00DF	Ô 00D4	Ň 0143	ň 0144	ň 0148	Š 0160	š 0161	Ŕ 0154	Ú 00DA	ŕ 0155	Ů 0170	ý 00FD	Ý 00DD	ţ 0163	´ 00B4
- 00AD	˘ 02DD	˘ 02DB	˘ 02C7	˘ 02D8	§ 00A7	÷ 00F7	˘ 00B8	◦ 00B0	˘ 00A8	˘ 02D9	ű 0171	Ř 0158	ř 0159	■ 25A0	◻ 00A0

formatos de archivos: representación de caracteres

UTF8 - UTF16 - UTF32 - UCS-2

- cada caracter es por varios bytes (eso es desventajoso)
- se puede representar cualquier caracter
- se puede incluir un prefijo para que quede clara cuál es la codificación
- en general se puede deducir la codificación mirando el texto
- menos propenso a errores
- extensible

ANSI

- cada caracter es representado por codificación directa, cada caracter ocupa 1 byte (8 bits).
- solo se puede representar caracteres de una única tabla por vez
- mirando el texto no hay manera de saber de qué tabla son los caracteres
- muy propenso a errores
- poco extensible

formatos de archivos: UTF-8 UTF-16

a = 65 = 100 0001
ñ = 241 = 1111 0001

CODIFICACIÓN CON UTF-8

ñ (Unicode) = U+00F1

= 0000 0000 F 1111 0001

110x xxxx

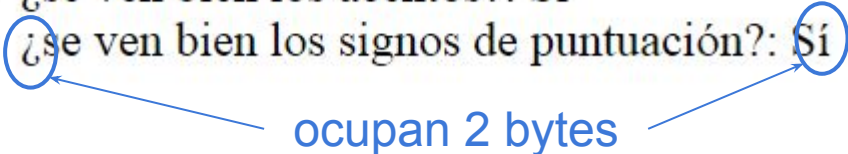
10xx xxxx

ñ (UTF-8) = 1100 0011 1011 0001 = C3B1

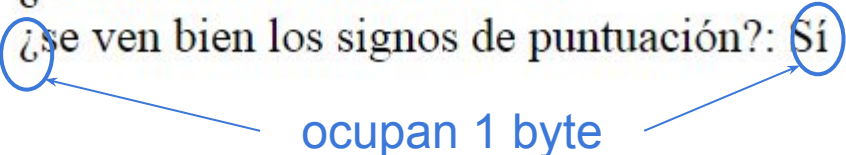
Rango de puntos UNICODE	Valor escalar	UTF-16	UTF-8
000000-00007F	00000000 0xxxxxxx	00000000 0xxxxxxx	0xxxxxxx
000080-0007FF	00000yyy yyxxxxxx	00000yyy yyxxxxxx	110yyyyy 10xxxxxx
000800-00FFFF	zzzzyyyy yyxxxxxx	zzzzyyyy yyxxxxxx	1110zzzz 10yyyyyy 10xxxxxx
010000-10FFFF	000uuuuu zzzzyyyy yyxxxxxx	110110ww wwzzzzyy 110111yy yyxxxxxx (www = uuuu - 1)	11110uuu 10uuzzzz 10yyyyyy 10xxxxxx

formatos de archivos: detectando errores

Prueba de UTF-8

- ¿se ven bien los acentos?: Sí
 - ¿se ven bien los signos de puntuación?: Sí
- ocupan 2 bytes
- 

Prueba de ANSI-1242

- ¿se ven bien los acentos?: Sí
 - ¿se ven bien los signos de puntuación?: Sí
- ocupan 1 byte
- 

formatos de archivos: detectando errores

Prueba de UTF-8

- ¿se ven bien los acentos?: SÃ
- ¿se ven bien los signos de puntuaciÃ³n?: SÃ

visto como ANSI-1242

Prueba de ANSI-1242

- ¿se ven bien los acentos?: S¿
- ¿se ven bien los signos de puntuaci¿n?: S¿

visto como UTF-8

Prueba de ANSI-1242

- ¿se ven bien los acentos?: Sí
- ¿se ven bien los signos de puntuación?: Sí

visto como ISO-8859-2