

# CODIFICACIÓN DE TEXTOS

## Fundamentos de Computación

Grado en Ingeniería en Tecnologías Industriales

Noviembre 2021

Universidad de Cantabria

## Texto llano

*Plain text*

```
Lorem ipsum dolor sit amet,
consectetuer adipiscing
elit. Aenean commodo ligula eget
dolor. Aenean massa. Cum sociis
natoque penatibus et magnis dis
parturient montes, nascetur
ridiculus mus. Donec quam felis,
ultricies nec, pellentesque eu,
pretium quis, sem. Nulla consequat
massa quis enim. Donec pede justo,
fringilla vel, aliquet nec,
vulputate eget, arcu. In enim
justo, rhoncus ut, imperdiet a,
venenatis vitae, justo. Nullam
dictum felis eu pede mollis
pretium. Integer tincidunt. Cras
dapibus. Vivamus elementum semper
nisi. Aenean vulputate eleifend
tellus. Aenean leo ligula,
porttitor eu, consequat vitae,
eleifend ac, enim. Aliquam lorem
ante, dapibus in, viverra quis,
feugiat a, tellus. Phasellus
viverra nulla ut metus varius
laoreet. Quisque rutrum. Aenean
```

.txt

## Texto con formato

DE FINIBVS BONORVM ET MALORVM

M. TVLLI CICERONIS

LIBER PRIMUS

NON ERAM NESCIUS, BRUTE, CUM, QUAE  
SUMMIS INGENIIS EXQUISITAQUE  
DOCTRINA PHILOSOPHI GRAECO SERMONE  
TRACTAVISSENT, EA LATINIS LITTERIS  
MANDAREMUS, FORE UT HIC NOSTER LABOR  
IN VARIAS REPREHENSIONES INCURRERET.  
NAM QUIBUSDAM, ET IIS QUIDEM NON  
ADMODUM INDOCTIS, TOTUM HOC  
DISPLICET PHILOSOPHARI. QUIDAM AUTEM  
NON TAM ID REPREHENDUNT, SI REMISSIUS  
AGATUR, SED TANTUM STUDIUM TAMQUE  
MULTAM OPERAM PONENDAM IN EO NON  
ARBITRANTUR. ERUNT ETIAM, ET II QUIDEM  
ERUDITI GRAECIS LITTERIS, CONTEMNENTES  
LATINAS, QUI SE DICANT IN GRAECIS  
LEGENDIS OPERAM MALLE CONSUMERE.  
POSTREMO ALIQUOS FUTUROS SUSPICOR,

.doc, .odt, .rtf, ...

# Editores y procesadores de texto





```

<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8">
</head>
<body bgcolor=#fffb05f>
<center>
<h1>EL TEOREMA DE EUCLIDES</h1>
<img src=./euclid.png
/><br><br>
<table width=50%>
<tr><td><h3 align=justify><b>La proposición 20 del libro IX de los
Elementos</i> de <a
href="http://en.wikipedia.org/wiki/Euclid"> Euclides</a> demuestra
que, para cualquier conjunto finito de números primos, siempre hay
un primo más.</b></h3></td></tr>
</table>
</center>
</body>
</html>

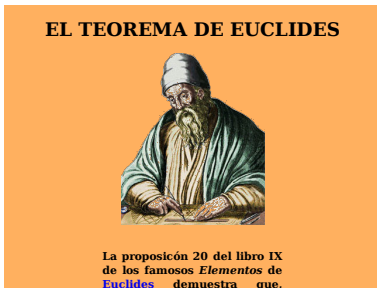
```

Documento fuente (.html, .htm)

Utiliza un *lenguaje de marcado* («markup language»).



Interpretado por un *navegador*



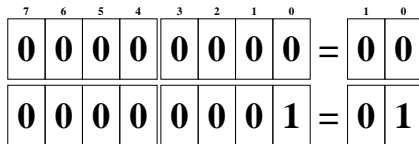
# Codificación de caracteres

Un byte por carácter  $\implies 2^8 = 256$  caracteres (desde 00 hasta FF)

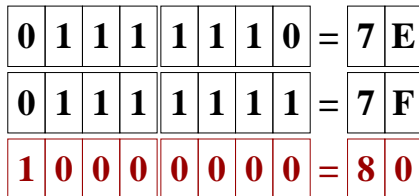
ASCII	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	RCK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2	SP	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6	`	a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

# Codificación de caracteres

Un byte por carácter  $\implies 2^8 = 256$



...



...



# Code page 1252 (Western)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
B	€		,	f	„	…	†	‡	^	‰	Š	<	œ		ž	
9		‘	’	“	”	•	–	—	˘	™	š	>	œ		ž	ÿ
A	NBSP	ı	ć	£	¤	¥	ı	§	¨	©	ª	«	¬	–	®	ˆ
B	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
C	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ



# ISO/IEC-8859-1 (Latin 1- Western European)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
A	␣	ı	ç	£	¤	¥	¦	§	¨	©	ª	«	¬	–	®	ˆ
B	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
C	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

# ISO/IEC-8859-2 (Latin 2- Central European)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	Š	š	Ž	ž	Ł	ł	Ś	ś	Ć	ć	Ę	ę	İ	ı	İ	ı
9	Š	š	Ž	ž	Ł	ł	Ś	ś	Ć	ć	Ę	ę	İ	ı	İ	ı
A	NBSP	Ą	ȧ	Ł	ł	Ś	ś	Ć	ć	Ę	ę	İ	ı	İ	ı	ı
B	°	ą	ȧ	ł	ł	ś	ś	ć	ć	ę	ę	ı	ı	ı	ı	ı
C	Ř	Á	Â	Ǻ	Ǻ	Í	Ć	Ç	Č	É	Ě	Ě	Í	Î	Ď	
D	Đ	Ń	Ň	Ó	Ô	Õ	Ö	×	Ř	Ů	Ú	Ů	Ü	Ý	Ť	ß
E	ř	á	â	ǻ	ǻ	í	ć	ç	č	é	ě	ě	í	î	ď	
F	đ	ń	ň	ó	ô	õ	ö	÷	ř	ů	ú	ů	ü	ý	ť	·

# ISO/IEC-8859-3 (Latin 3- South European)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
A	NBSP	Ħ	˘	£	¤		Ħ	§	¨	ı	Ş	Č	Ĵ	-		Ž
B	◦	Ħ	²	³	´	µ	ñ	·	¸	ı	ş	č	ĵ	½		ž
C	À	Á	Â		Ä	Č	Ĉ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D		Ñ	Ò	Ó	Ô	Ĝ	Ö	×	Ĝ	Ù	Ú	Û	Ü	Ů	Š	ß
E	à	á	â		ä	č	ĉ	ç	è	é	ê	ë	ì	í	î	ï
F		ñ	ò	ó	ô	ğ	ö	÷	ğ	ù	ú	û	ü	ů	š	·

# ISO/IEC-8859-4 (Latin 4- North European)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	Š	š	Ž	ž	Č	č	Š	š	Ž	ž	Č	č	Š	š	Ž	ž
9	Š	š	Ž	ž	Č	č	Š	š	Ž	ž	Č	č	Š	š	Ž	ž
A	NBSP	Ą	ą	Ŕ	ŕ	Ī	Ĳ	Š	š	Ė	Ģ	Ŧ	-	Ž	-	
B	°	ą	.	ŗ	ı	ı	˘	.	š	ė	ģ	ŧ	Ŧ	ž	ŋ	
C	Ā	Ā	Ā	Ā	Ā	Ā	Æ	Į	Č	É	Ē	Ē	É	Í	Ī	
D	Ð	Ņ	Ō	Ŕ	Ô	Ö	×	Ø	Ū	Ú	Û	Ü	Ū	Ū	ß	
E	ā	ā	ā	ā	ā	ā	æ	į	č	é	ē	ē	é	í	ī	
F	đ	ņ	ō	ŗ	ô	ö	÷	ø	ų	ú	û	ü	ū	ū	·	

# ISO/IEC-8859-5 (Latin/Cyrillic)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
A	NBSP	Ё	Ђ	Ѓ	Є	Ѕ	І	Ї	Ј	Љ	Њ	Ћ	Ќ	–	Ў	Ц
B	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
C	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
D	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
E	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
F	№	ё	ђ	ѓ	є	ѕ	і	ї	ј	љ	њ	ћ	ќ	–	ў	ц

# ISO/IEC-8859-6 (Latin/Arabic)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
9	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
A	NBSP				□								،	□		
B												:				؟
C		ء	آ	أ	ؤ	إ	ئ	ا	ب	ة	ت	ث	ج	ح	خ	د
D	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	غ					
E	.	ف	ق	ك	ل	م	ن	ه	و	ى	ي	ء	ء	ء	ء	ء
F	ء	ء	ء													

# ISO/IEC-8859-7(Latin/Greek)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
9	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
A	NBSP	'	'	£	€	đ	ı	§	”	©	.	«	¬	-		-
B	°	±	²	³	´	“	À	·	Ɛ	Ɔ	Ɔ	»	Ɔ	½	Ɔ	ŉ
C	í	A	B	Γ	Δ	E	Z	H	θ	I	K	Λ	M	N	Ξ	Ο
D	Π	P		Σ	T	Υ	Φ	X	Ψ	Ω	İ	ÿ	ά	έ	ή	ί
E	ú	α	β	γ	δ	ε	ζ	η	θ	ι	κ	λ	μ	ν	ξ	ο
F	π	ρ	ς	σ	τ	υ	φ	χ	ψ	ω	ï	ü	ó	ú	ώ	

# ISO/IEC-8859-8 (Latin/Hebrew)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
A	NBSP	☐	¢	£	¤	¥	¦	§	¨	©	×	«	¬	–	®	ˆ
B	°	±	²	³	´	µ	¶	·	¸	¹	÷	»	¼	½	¾	☐
C	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
D	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	–
E	א	ב	ג	ד	ה	ו	ז	ח	ט	י	ך	כ	ל	מ	נ	ן
F	נ	ס	ע	ף	פ	ץ	צ	ק	ר	ש	ת	☐	☐	☐	☐	☐



# ISO/IEC-8859-9(Latin 5- Turkish)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
A	NBSP	ı	ç	£	¤	¥	¦	§	¨	©	ª	«	¬	–	®	ˆ
B	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
C	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D	Ğ	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	İ	Ş	ß
E	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F	ğ	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ı	ş	ÿ

# ISO/IEC-8859-10 (Latin 6- Nordic)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø
9	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø	ø
A	NBSP	Ą	Ĕ	Ģ	Ī	Ī	Ķ	Š	Ł	Ð	Š	Ŧ	Ž	-	Ū	Ų
B	°	ą	ē	ģ	ī	ī	ķ	·	ł	đ	š	ŧ	ž	-	ū	ų
C	Ā	Ā	Ā	Ā	Ā	Ā	Æ	Į	Č	É	Ē	Ē	Ē	Í	Ī	Ī
D	Ð	Ŋ	Ō	Ó	Ô	Õ	Ö	Û	Ø	Ų	Ú	Û	Ü	Ý	Þ	ß
E	ā	ā	ā	ā	ā	ā	æ	į	č	é	ē	ē	ē	í	ī	ī
F	ð	ŋ	ō	ó	ô	õ	ö	ū	ø	ų	ú	û	ü	ý	þ	ÿ

# ISO/IEC-8859-11(Latin/Thai)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
9	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
A	NBSP	ภ	ข	ช	ค	ค	ฆ	ง	จ	ฉ	ช	ช	ฌ	ญ	ฎ	ฏ
B	ฐ	ฑ	ฒ	ณ	ด	ด	ถ	ท	ธ	น	บ	ป	ผ	ฝ	พ	ฟ
C	ภ	ม	ย	ร	ฤ	ล	ฎ	ว	ศ	ษ	ส	ห	ฬ	อ	ฮ	า
D	ะ	ั	า	ำ	ิ	ี	ึ	ุ	ุ	ุ						๕
E	เ	แ	โ	ใ	ไ	ำ	ำ	ั	ิ	ึ	ุ	ุ	ุ	ุ	ุ	ุ
F	๐	๑	๒	๓	๔	๕	๖	๗	๘	๙	๐	๑				

# ISO/IEC-8859-13(Latin 7- Baltic Rim)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
A	NBSP	"	¢	£	¤	„	¡	§	ø	©	Ŕ	«	¬	-	®	Æ
B	°	±	²	³	“	µ	¶	·	ø	¹	ŗ	»	¼	½	¾	æ
C	Ą	Į	Ā	Ć	Ä	Å	Ę	Ē	Č	É	Ž	È	Ğ	Ķ	Ī	Ļ
D	Š	Ń	Ņ	Ó	Õ	Ö	Ö	×	Ų	Ł	Ś	Ū	Ü	Ž	Ž	ß
E	ą	į	ā	ć	ä	å	ę	ē	č	é	ž	è	ğ	ķ	ī	ļ
F	š	ń	ņ	ó	õ	ö	ö	÷	ų	ł	ś	ū	ü	ž	ž	'

# ISO/IEC-8859-14 (Latin 8- Celtic)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
A	NBSP	È	é	£	Ĉ	ĉ	Ď	Š	Ŵ	©	Ŷ	à	ÿ	-	®	Ÿ
B	ƒ	ƒ	Ġ	ġ	Ĺ	ĺ	Ŧ	Ɔ	ŵ	Ɔ	ŵ	Š	ÿ	Ŵ	ŵ	š
C	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D	Ŵ	Ŷ	Ò	Ó	Ô	Õ	Ö	Ŧ	Ø	Ù	Ú	Û	Ü	Ý	Ÿ	ß
E	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F	ŵ	Ŷ	ò	ó	ô	õ	ö	ŧ	ø	ù	ú	û	ü	ý	ÿ	ÿ

# ISO/IEC-8859-15 (Latin 9- Revision Latin -1)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
A	NBSP	ı	ç	£	€	¥	Š	š	š	©	ª	«	¬	-	®	-
B	°	±	²	³	ž	µ	¶	·	ž	¹	º	»	œ	œ	Ÿ	ı
C	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

# ISO/IEC-8859-16 (Latin 10- South-Eastern European)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
A	NBSP	Ą	ą	Ł	ł	„	Š	š	Š	š	€	«	Ž	-	ž	Ž
B	°	±	Č	č	Ž	ž	ŕ	.	ž	č	ŕ	»	œ	œ	Ÿ	z
C	À	Á	Â	Ă	Ä	Ć	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	Ś	Ũ	Ù	Ú	Û	Ü	Œ	Ŧ	ß
E	à	á	â	ă	ä	ć	æ	ç	è	é	ê	ë	ì	í	î	ï
F	đ	ń	ò	ó	ô	õ	ö	ś	ű	ù	ú	û	ü	ę	ț	ÿ

## Universal Character Set (UCS)

Repertorio de caracteres que incluye letras de multitud de idiomas, símbolos, ideogramas, etc.

Se divide en *planos*, cada uno de ellos con  $65\,536 = 2^{16} = (2^4)^4 = (2^8)^2$  caracteres (*code points*).

Un carácter dentro de un plano queda determinado por 2 bytes:  
00 00  $\longrightarrow$  FF FF.

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



```
>>> chr(1092)
'ϕ'
>>> chr(10920)
'◄'
>>> chr(920)
'θ'
>>> chr(960)
'π'
>>> ord('E')
69
>>> int('0x1f61',16)
8033
>>> chr(8033)
'ώ'
>>> int('0x1f601',16)
128513
>>> chr(128513)
'😄'
>>> ord('◄')
10920
```

## Basic Multilingual Plane (BMP)

El primero de los planos

La codificación de caracteres UCS-2 (obsoleta) identifica cada carácter de este plano utilizando sencillamente su posición.

Se utilizan, por tanto, 2 bytes por cada carácter:

00 73:	s	Quando el primer byte es 00, el segundo coincide con la codificación ISO-8859-1.
00 C6:	Æ	
30 6E:	の	

## Universal Character Set (UCS)

No solo está el «Basic Multilingual Plane». Hay 17 planos, que se etiquetan con los prefijos:

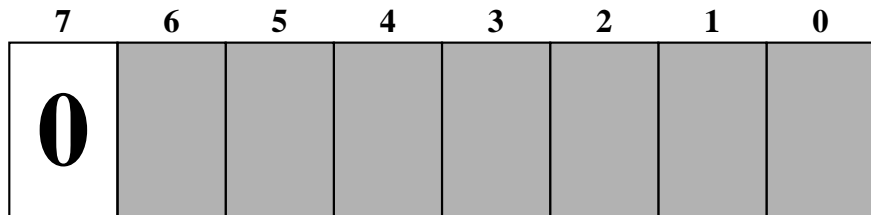
00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 0A, 0B, 0C, 0D, 0E, 0F y 10.

Esto da un total de  $17 \times 2^{16} = 1\,114\,112$  *code points*.

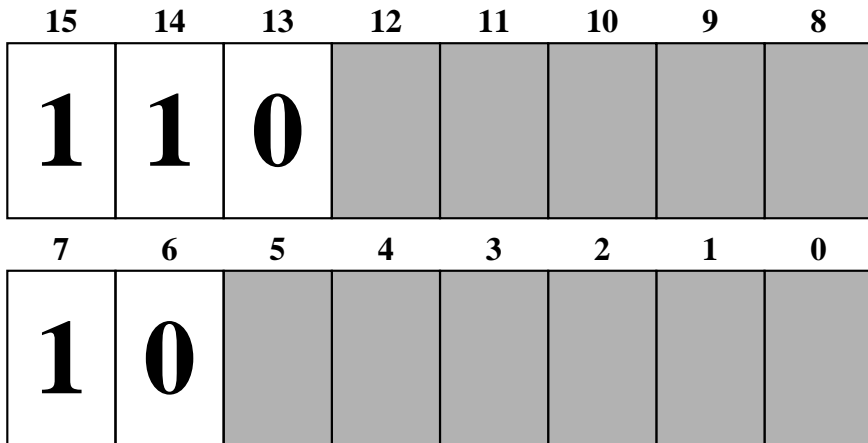
## UCS Transformation Format — 8 bit

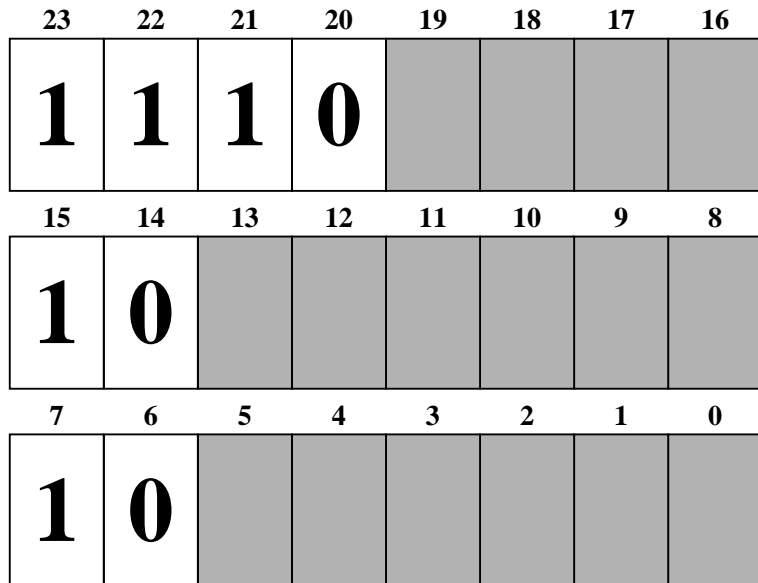
Es una codificación de **longitud variable**:

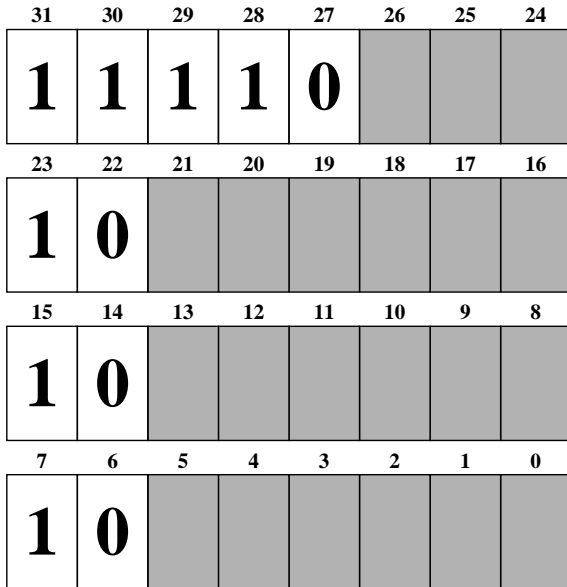
Caracteres del rango	bits del <i>code point</i>	bytes UTF-8
0 - 7F (ASCII)	1-7	1
80 - 07 FF	8-11	2
08 00 - FF FF	12-16	3
01 00 00 - 1F FF FF	17-21	4



Coincide con la codificación ISO-8859.









J

ASCII

0	1	0	0	1	0	1	0
---	---	---	---	---	---	---	---

4	A
---	---

ISO-8859

0	1	0	0	1	0	1	0
---	---	---	---	---	---	---	---

4	A
---	---

UCS-2

0	0	0	0	0	0	0	0
0	1	0	0	1	0	1	0

0	0
4	A

UTF-8

0	1	0	0	1	0	1	0
---	---	---	---	---	---	---	---

4	A
---	---

o

ASCII

NO

ISO-8859-1

1	0	1	1	1	0	1	0
---	---	---	---	---	---	---	---

B	A
---	---

UCS-2

0	0	0	0	0	0	0	0
1	0	1	1	1	0	1	0

0	0
B	A

UTF-8

1	1	0	0	0	0	1	0
1	0	1	1	1	0	1	0

C	2
B	A

ñ

ASCII

NO

ISO-8859-1

1	1	1	1	0	0	0	1
---	---	---	---	---	---	---	---

F	1
---	---

UCS-2

0	0	0	0	0	0	0	0
1	1	1	1	0	0	0	1

0	0
F	1

UTF-8

1	1	0	0	0	0	1	1
1	0	1	1	0	0	0	1

C	3
B	1

$\pi$

ASCII

NO

ISO-8859-7

1	1	1	1	0	0	0	0
---	---	---	---	---	---	---	---

F	0
---	---

UCS-2

0	0	0	0	0	0	1	1
1	1	0	0	0	0	0	0

0	3
C	0

UTF-8

1	1	0	0	1	1	1	1
1	0	0	0	0	0	0	0

C	F
8	0

か

ASCII

NO

ISO-8859

NO

UCS-2

0	0	1	1	0	0	0	0
0	1	0	0	1	0	1	1

3	0
4	B

UTF-8

1	1	1	0	0	0	1	1
1	0	0	0	0	0	0	1
1	0	0	0	1	0	1	1

E	3
8	1
8	B