

# Extraction d'information

## Cours 3

---

Nassim ZELLAL

# Table ASCII

## ASCII TABLE

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]



## Analyse avec Winhex - ASCII

# Analyse avec Winhex - ASCII

0	1	2	3	4	5	6	7	8	9	A	B	
63	6F	65	75	72								coeur
<b>5 octets</b>												

# Caractère spécial « oe »

1 cœur

**Créer un fichier en Windows-1252  
puis y écrire cœur.**

**Enregistrer ce fichier avec  
l'encodage « Windows-1252 ».**

# Analyse avec Winhex - Windows 1252

0	1	2	3	4	5	6	7	8	9	A	B	
63	9C	75	72									c ur

**4 octets**

# Table Windows-1252

Windows-1252 (CP1252)

	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	xA	xB	xC	xD	xE	xF
0x	<u>NUL</u>	<u>SOH</u>	<u>STX</u>	<u>ETX</u>	<u>EOT</u>	<u>ENQ</u>	<u>ACK</u>	<u>BEL</u>	<u>BS</u>	<u>HT</u>	<u>LF</u>	<u>VT</u>	<u>FF</u>	<u>CR</u>	<u>SO</u>	<u>SI</u>
1x	<u>DLE</u>	<u>DC1</u>	<u>DC2</u>	<u>DC3</u>	<u>DC4</u>	<u>NAK</u>	<u>SYN</u>	<u>ETB</u>	<u>CAN</u>	<u>EM</u>	<u>SUB</u>	<u>ESC</u>	<u>FS</u>	<u>GS</u>	<u>RS</u>	<u>US</u>
2x	<u>SP</u>	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4x	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5x	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6x	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7x	p	q	r	s	t	u	v	w	x	y	z	{		}	~	<u>DEL</u>
8x	€		,	f	„	...	†	‡	^	‰	Š	‹	Œ		Ž	
9x		‘	’	“	”	•	—	—	~	™	š	›	œ		ž	ÿ
Ax	<u>NBSP</u>	ı	¢	£	¤	¥	¦	§	¨	©	ª	«	¬		®	¯
Bx	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
Cx	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
Dx	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
Ex	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
Fx	ø	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

# Analyse avec Winhex - ISO-8859-1 (Latin-1)

0	1	2	3	4	5	6	7	8	9	A	B	
63	6F	75	72									our

L'enregistrer avec l'encodage « ISO-8859-1 »



# Table ISO-8859-1 (Latin-1)

## ISO/CEI 8859-1

	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	xA	xB	xC	xD	xE	xF
0x	<i>positions inutilisées</i>															
1x																
2x	SP	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4x	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5x	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6x	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7x	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8x	<i>positions inutilisées</i>															
9x																
Ax	NBSP	í	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	-	®	¯
Bx	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
Cx	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
Dx	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
Ex	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
Fx	ø	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

# UNICODE

- Unicode est un mécanisme universel de codage des caractères.
- Parmi les encodages du standard UNICODE, nous avons :
  - UTF-8 sans BOM ;
  - UTF-8 avec BOM (séquence BOM = EF BB BF);
  - UTF-16 Big Endian (séquence BOM = FE FF);
  - UTF-16 Little Endian (séquence BOM = FF FE);
  - UTF-32 Big Endian (séquence BOM = 00 00 FE FF);
  - UTF-32 Little Endian (séquence BOM = FF FE 00 00);
  - BOM = Byte Order Mark = indicateur d'ordre des octets (cas de l'UTF-16) + indique l'utilisation d'un encodage UNICODE (cas de l'UTF-8 et de l'UTF-16).

# UNICODE - UTF-8 sans BOM

table

text file

length : 5 lines : 1

Ln : 1 Col : 6 Sel : 0 | 0

Windows (CR LF)

UTF-8

## Analyse avec Winhex - UTF-8 sans BOM

[illegible]

# UNICODE - UTF-8 avec BOM

table

text file      length : 5    lines : 1      ln : 1    Col : 6    Sel : 0 | 0      Windows (CR LF)      **UTF-8-BOM**

## Analyse avec Winhex - UTF-8 avec BOM

# Analyse avec Winhex - UTF-8 avec BOM

# Caractères du BOM UTF-8 dans la table ISO-8859-1

ISO/CEI 8859-1

	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	xA	xB	xC	xD	xE	xF
0x	<i>positions inutilisées</i>															
1x																
2x	SP	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4x	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5x	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6x	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7x	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8x	<i>positions inutilisées</i>															
9x																
Ax	NBSP	ı	ç	£	¤	¥	¦	§	¨	©	ª	«	¬	-	®	¯
Bx	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
Cx	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
Dx	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
Ex	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
Fx	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

# Unicode - UTF-8 avec BOM

école



# Analyse avec Winhex - UTF-8 avec BOM

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
EF	BB	BF	C3	A9	63	6F	6C	65								i»lÅ@cole

**Au total = 9 octets**

# UNICODE - UTF-16

- L'encodage UTF-16 (UCS-2 : UCS = Universal Coded Character Set - ISO/IEC 10646 - encodage sur deux octets) possède un **BOM** (Byte Order Mark : Indicateur d'Ordre des Octets).
- En UTF-16, ce BOM représente une suite de caractères utilisée au début d'un fichier pour marquer l'**endianness** d'une chaîne de caractères UNICODE.
- FE (1<sup>er</sup> octet) FF (2<sup>ème</sup> octet) : Big-Endian (BE) → þ ÿ
- FF (1<sup>er</sup> octet) FE (2<sup>ème</sup> octet) : Little-Endian (LE) → ÿ þ

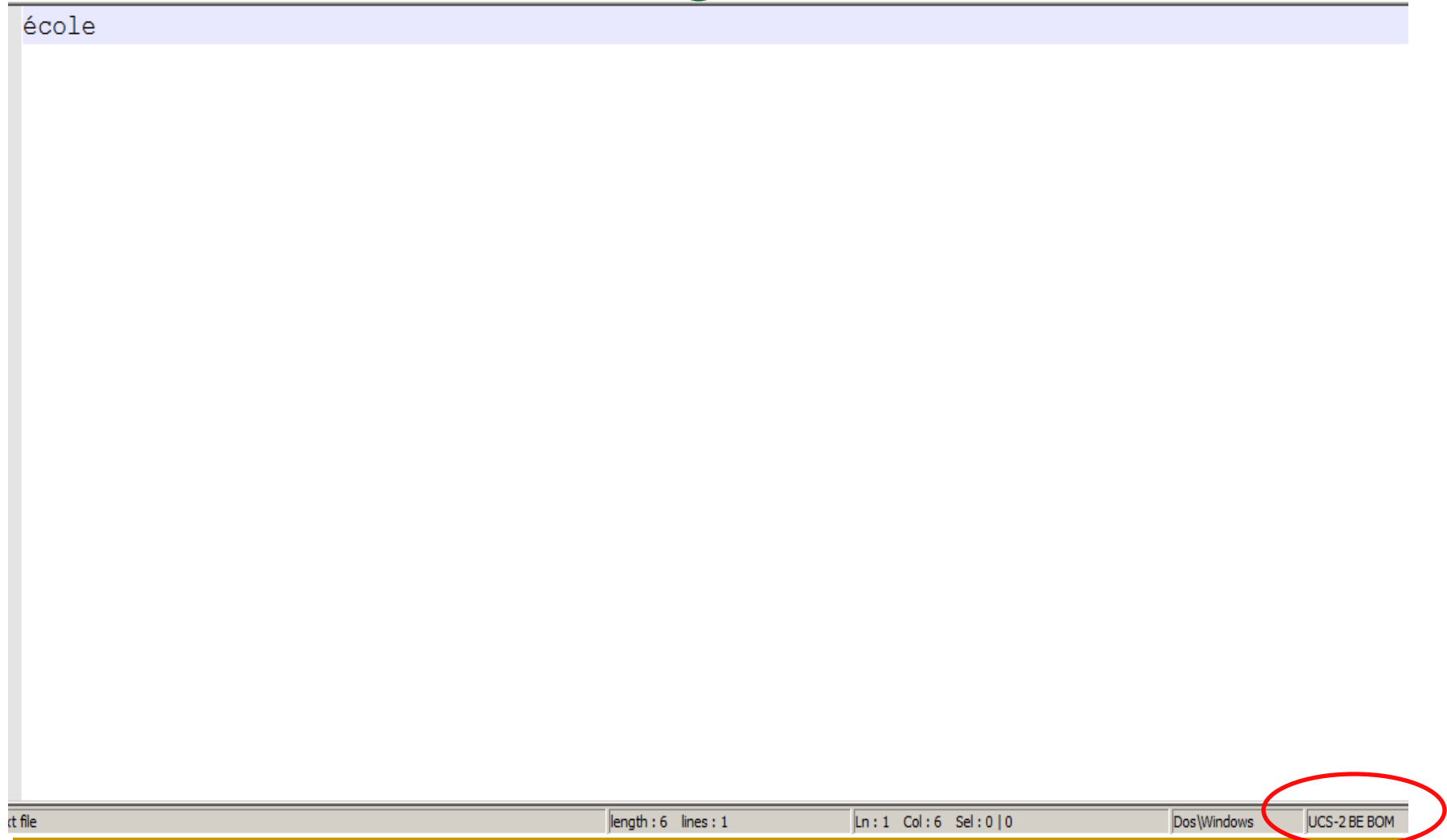
# Caractères du BOM UTF-16 dans la table ISO-8859-1

ISO/CEI 8859-1																
	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	xA	xB	xC	xD	xE	xF
0x	<i>positions inutilisées</i>															
1x																
2x	SP	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4x	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5x	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6x	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7x	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8x	<i>positions inutilisées</i>															
9x																
Ax	NBSP	ı	ø	£	¤	¥	¦	§	¨	©	ª	«	¬	­	®	¯
Bx	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
Cx	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
Dx	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
Ex	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
Fx	ø	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

# UNICODE - ordre des octets en UTF-16

- Prenons l'exemple du mot « école » (5 caractères).
- Regardons sur l'éditeur hexadécimal WinHex l'ordre des octets de ce mot.
- Le premier fichier est encodé en UTF-16BE BOM et le second en UTF-16LE BOM.

# Ordre des octets en UTF-16 - Big Endian - BOM



# Ordre des octets en UTF-16 - Big Endian - BOM

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
FE	FF	00	E9	00	63	00	6F	00	6C	00	65					þý.é.c.o.l.e

# Ordre des octets en UTF-16 - Little Endian - BOM

école

file

length : 6 lines : 1

Ln : 1 Col : 6 Sel : 0 | 0

Dos\Windows

UCS-2 LE BOM

# Ordre des octets en UTF-16 - Little Endian - BOM

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
FF	FE	E9	00	63	00	6F	00	6C	00	65	00					ýpé.c.o.l.e.

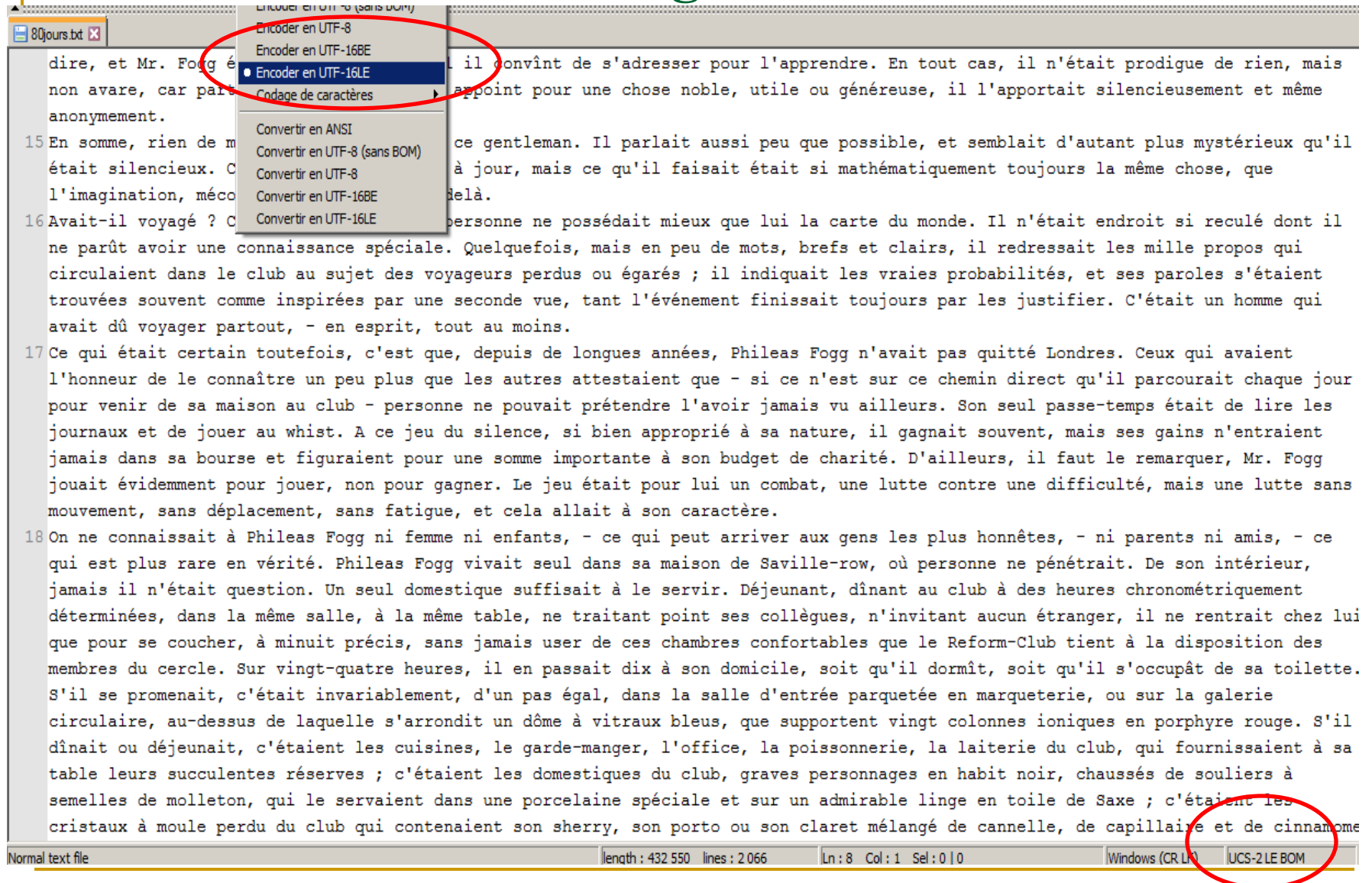


# Gestion de l'encodage d'un texte - Unix

```
80jours.txt x
dire, et Mr. Fogg était le dernier auquel il convînt de s'adresser pour l'apprendre. En tout cas, il n'était prodigue de rien, mais non avare, car partout où il manquait un appoint pour une chose noble, utile ou généreuse, il l'apportait silencieusement et même anonymement.
15 En somme, rien de moins communicatif que ce gentleman. Il parlait aussi peu que possible, et semblait d'autant plus mystérieux qu'il était silencieux. Cependant sa vie était à jour, mais ce qu'il faisait était si mathématiquement toujours la même chose, que l'imagination, mécontente, cherchait au-delà.
16 Avait-il voyagé ? C'était probable, car personne ne possédait mieux que lui la carte du monde. Il n'était endroit si reculé dont il ne parût avoir une connaissance spéciale. Quelquefois, mais en peu de mots, brefs et clairs, il redressait les mille propos qui circulaient dans le club au sujet des voyageurs perdus ou égarés ; il indiquait les vraies probabilités, et ses paroles s'étaient trouvées souvent comme inspirées par une seconde vue, tant l'événement finissait toujours par les justifier. C'était un homme qui avait dû voyager partout, - en esprit, tout au moins.
17 Ce qui était certain toutefois, c'est que, depuis de longues années, Phileas Fogg n'avait pas quitté Londres. Ceux qui avaient l'honneur de le connaître un peu plus que les autres attestaient que - si ce n'est sur ce chemin direct qu'il parcourait chaque jour pour venir de sa maison au club - personne ne pouvait prétendre l'avoir jamais vu ailleurs. Son seul passe-temps était de lire les journaux et de jouer au whist. A ce jeu du silence, si bien approprié à sa nature, il gagnait souvent, mais ses gains n'entraient jamais dans sa bourse et figuraient pour une somme importante à son budget de charité. D'ailleurs, il faut le remarquer, Mr. Fogg jouait évidemment pour jouer, non pour gagner. Le jeu était pour lui un combat, une lutte contre une difficulté, mais une lutte sans mouvement, sans déplacement, sans fatigue, et cela allait à son caractère.
18 On ne connaissait à Phileas Fogg ni femme ni enfants, - ce qui peut arriver aux gens les plus honnêtes, - ni parents ni amis, - ce qui est plus rare en vérité. Phileas Fogg vivait seul dans sa maison de Saville-row, où personne ne pénétrait. De son intérieur, jamais il n'était question. Un seul domestique suffisait à le servir. Déjeunant, dînant au club à des heures chronométriquement déterminées, dans la même salle, à la même table, ne traitant point ses collègues, n'invitant aucun étranger, il ne rentrait chez lui que pour se coucher, à minuit précis, sans jamais user de ces chambres confortables que le Reform-Club tient à la disposition des membres du cercle. Sur vingt-quatre heures, il en passait dix à son domicile, soit qu'il dormît, soit qu'il s'occupât de sa toilette. S'il se promenait, c'était invariablement, d'un pas égal, dans la salle d'entrée parquetée en marqueterie, ou sur la galerie circulaire, au-dessus de laquelle s'arrondit un dôme à vitraux bleus, que supportent vingt colonnes ioniques en porphyre rouge. S'il dînait ou déjeunait, c'étaient les cuisines, le garde-manger, l'office, la poissonnerie, la laiterie du club, qui fournissaient à sa table leurs succulentes réserves ; c'étaient les domestiques du club, graves personnages en habit noir, chaussés de souliers à semelles de molleton, qui le servaient dans une porcelaine spéciale et sur un admirable linge en toile de Saxe ; c'étaient les cristaux à moule perdu du club qui contenaient son sherry, son porto ou son claret mélangé de cannelle, de capillaire et de cinnamome
Normal text file                               length: 432 550   lines: 2 066                               Ln: 8   Col: 1   Sel: 0 | 0                               Windows (CR LF)                               UCS-2 LE BOM
```

**UCS = Universal Coded Character  
Set - ISO/IEC 10646**

# Gestion de l'encodage d'un texte - Unix



dire, et Mr. Fogg é non avare, car par anonymement.

15 En somme, rien de m était silencieux. C l'imagination, méco

16 Avait-il voyagé ? C personne ne possédait mieux que lui la carte du monde. Il n'était endroit si reculé dont il ne parût avoir une connaissance spéciale. Quelquefois, mais en peu de mots, brefs et clairs, il redressait les mille propos qui circulaient dans le club au sujet des voyageurs perdus ou égarés ; il indiquait les vraies probabilités, et ses paroles s'étaient trouvées souvent comme inspirées par une seconde vue, tant l'événement finissait toujours par les justifier. C'était un homme qui avait dû voyager partout, - en esprit, tout au moins.

17 Ce qui était certain toutefois, c'est que, depuis de longues années, Phileas Fogg n'avait pas quitté Londres. Ceux qui avaient l'honneur de le connaître un peu plus que les autres attestaient que - si ce n'est sur ce chemin direct qu'il parcourait chaque jour pour venir de sa maison au club - personne ne pouvait prétendre l'avoir jamais vu ailleurs. Son seul passe-temps était de lire les journaux et de jouer au whist. A ce jeu du silence, si bien approprié à sa nature, il gagnait souvent, mais ses gains n'entraient jamais dans sa bourse et figuraient pour une somme importante à son budget de charité. D'ailleurs, il faut le remarquer, Mr. Fogg jouait évidemment pour jouer, non pour gagner. Le jeu était pour lui un combat, une lutte contre une difficulté, mais une lutte sans mouvement, sans déplacement, sans fatigue, et cela allait à son caractère.

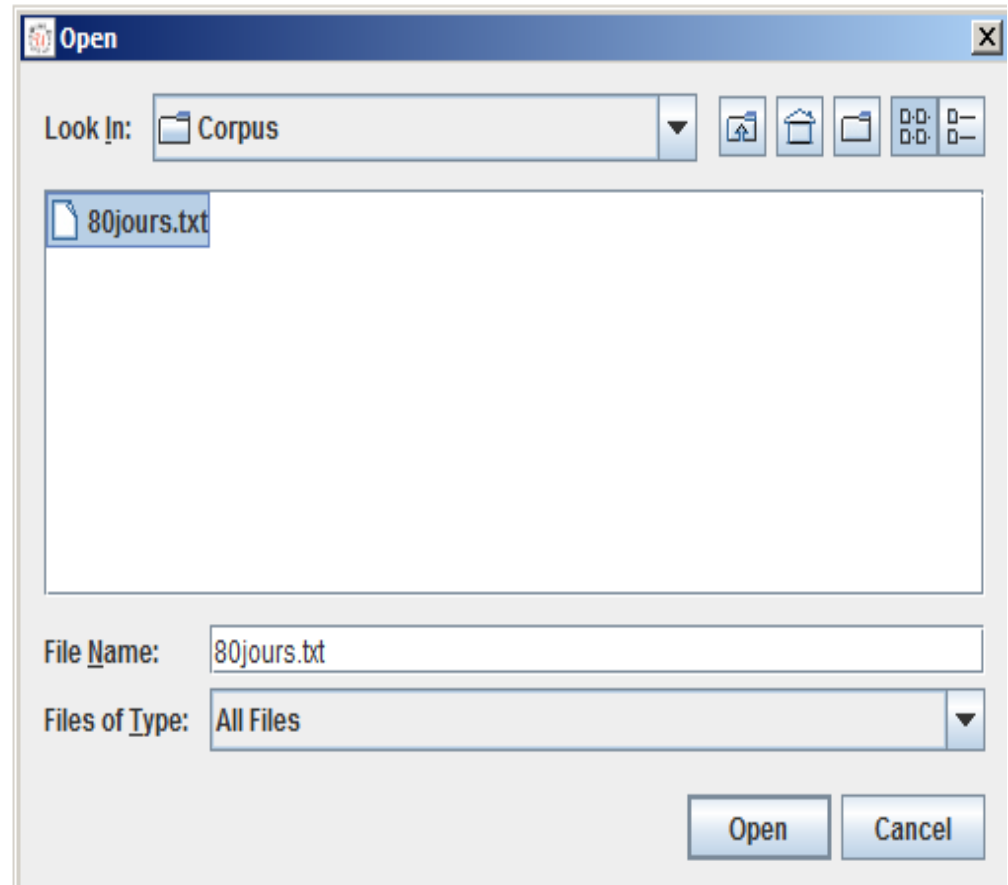
18 On ne connaissait à Phileas Fogg ni femme ni enfants, - ce qui peut arriver aux gens les plus honnêtes, - ni parents ni amis, - ce qui est plus rare en vérité. Phileas Fogg vivait seul dans sa maison de Saville-row, où personne ne pénétrait. De son intérieur, jamais il n'était question. Un seul domestique suffisait à le servir. Déjeunant, dînant au club à des heures chronométriquement déterminées, dans la même salle, à la même table, ne traitant point ses collègues, n'invitant aucun étranger, il ne rentrait chez lui que pour se coucher, à minuit précis, sans jamais user de ces chambres confortables que le Reform-Club tient à la disposition des membres du cercle. Sur vingt-quatre heures, il en passait dix à son domicile, soit qu'il dormît, soit qu'il s'occupât de sa toilette. S'il se promenait, c'était invariablement, d'un pas égal, dans la salle d'entrée parquetée en marqueterie, ou sur la galerie circulaire, au-dessus de laquelle s'arrondit un dôme à vitraux bleus, que supportent vingt colonnes ioniques en porphyre rouge. S'il dînait ou déjeunait, c'étaient les cuisines, le garde-manger, l'office, la poissonnerie, la laiterie du club, qui fournissaient à sa table leurs succulentes réserves ; c'étaient les domestiques du club, graves personnages en habit noir, chaussés de souliers à semelles de molleton, qui le servaient dans une porcelaine spéciale et sur un admirable linge en toile de Saxe ; c'étaient les cristaux à moule perdu du club qui contenaient son sherry, son porto ou son claret mélangé de cannelle, de capillaire et de cinnamome

Normal text file length: 432 550 lines: 2 066 Ln: 8 Col: 1 Sel: 0 | 0 Windows (CR LF) UCS-2 LE BOM

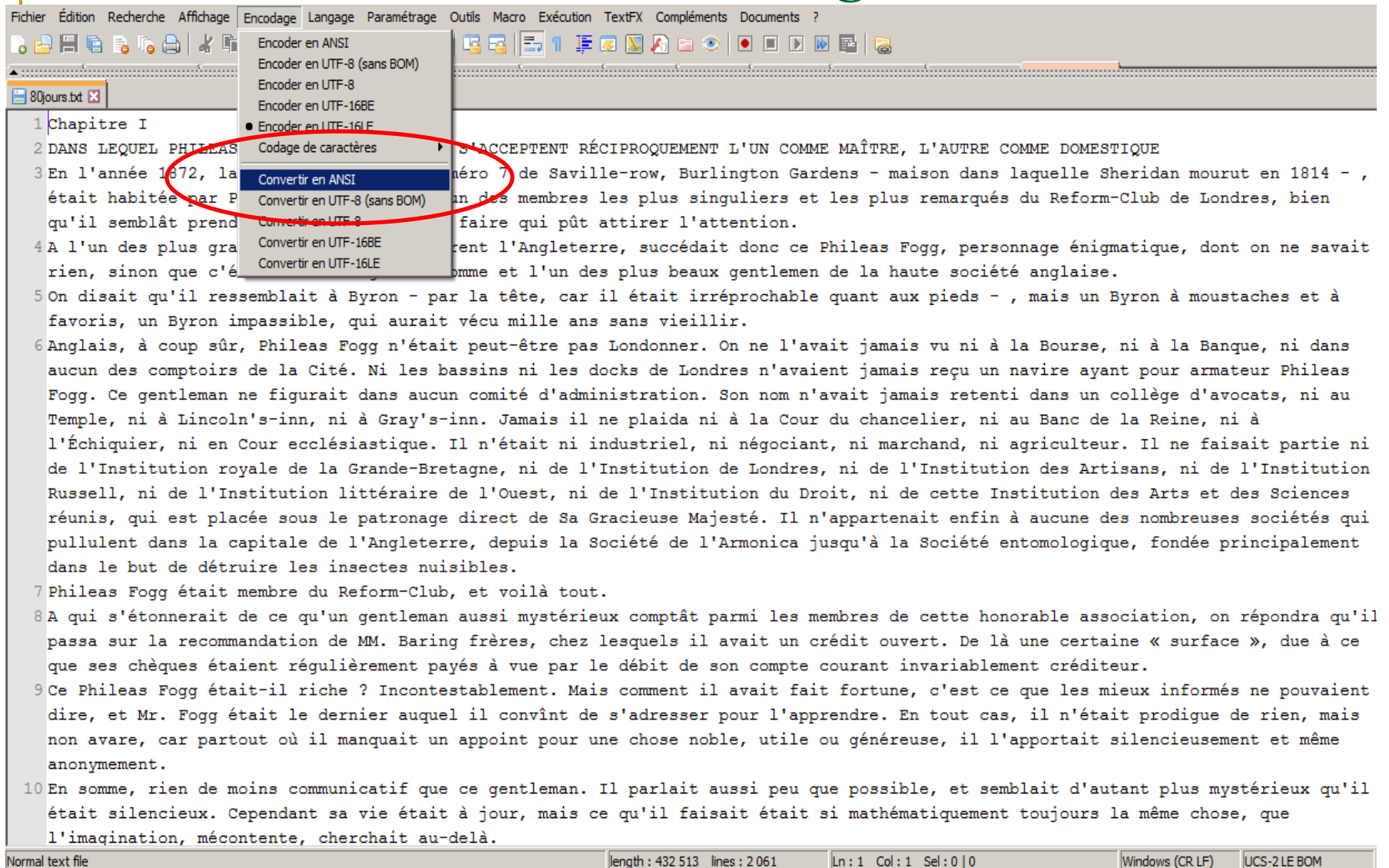
**UTF (UCS Transformation Format) -16 LE (Little Endian) BOM (Byte Order Mark) est l'encodage utilisé par défaut par Unitex. Cet encodage appartient au standard UNICODE.**

**Unicode est un mécanisme universel de codage des caractères.**

**Si on convertit cet encodage vers ANSI (voir slide suivant), Unitex vous propose de faire une conversion d'encodage.**



# Conversion de l'encodage d'un texte



The screenshot shows a text editor window with a menu bar at the top: Fichier, Édition, Recherche, Affichage, Encodage, Langage, Paramétrage, Outils, Macro, Exécution, TextFX, Compléments, Documents ?.

The 'Encodage' menu is open, showing the following options:

- Encoder en ANSI
- Encoder en UTF-8 (sans BOM)
- Encoder en UTF-8
- Encoder en UTF-16BE
- Encoder en UTF-16LE
- Codage de caractères
- Convertir en ANSI**
- Convertir en UTF-8 (sans BOM)
- Convertir en UTF-8
- Convertir en UTF-16BE
- Convertir en UTF-16LE

The text in the editor is a French translation of a chapter from 'Les Mystères de la rue Marmaduke' by Wilkie Collins. The text is displayed in a monospaced font. The status bar at the bottom shows: Normal text file, length: 432 513 lines: 2 061, Ln: 1 Col: 1 Sel: 0 | 0, Windows (CR LF), UCS-2 LE BOM.

1 Chapitre I

2 DANS LEQUEL PHILEAS

3 En l'année 1872, la

4 A l'un des plus gra

5 On disait qu'il ressemblait à Byron - par la tête, car il était irréprochable quant aux pieds - , mais un Byron à moustaches et à

6 Anglais, à coup sûr, Phileas Fogg n'était peut-être pas Londonner. On ne l'avait jamais vu ni à la Bourse, ni à la Banque, ni dans

7 Phileas Fogg était membre du Reform-Club, et voilà tout.

8 A qui s'étonnerait de ce qu'un gentleman aussi mystérieux comptât parmi les membres de cette honorable association, on répondra qu'il

9 Ce Phileas Fogg était-il riche ? Incontestablement. Mais comment il avait fait fortune, c'est ce que les mieux informés ne pouvaient

10 En somme, rien de moins communicatif que ce gentleman. Il parlait aussi peu que possible, et semblait d'autant plus mystérieux qu'il

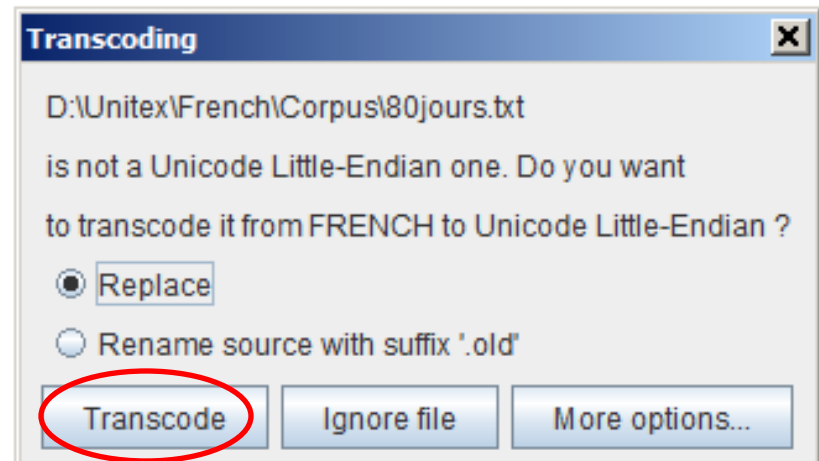
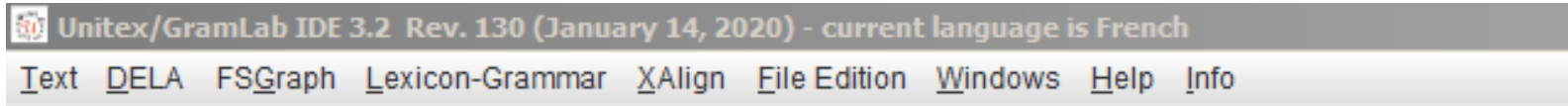


# Conversion de l'encodage d'un texte

```
80jours.txt x
1 Chapitre I
2 DANS LEQUEL PHILEAS FOGG ET PASSEPARTOUT S'ACCEPTENT RÉCIPROQUEMENT L'UN COMME MAÎTRE, L'AUTRE COMME DOMESTIQUE
3 En l'année 1872, la maison portant le numéro 7 de Saville-row, Burlington Gardens - maison dans laquelle Sheridan mourut en 1814 - , était habitée par Phileas
  Fogg, esq., l'un des membres les plus singuliers et les plus remarqués du Reform-Club de Londres, bien qu'il semblât prendre à tâche de ne rien faire qui pût
  attirer l'attention.
4 A l'un des plus grands orateurs qui honorent l'Angleterre, succédait donc ce Phileas Fogg, personnage énigmatique, dont on ne savait rien, sinon que c'était un
  fort galant homme et l'un des plus beaux gentlemen de la haute société anglaise.
5 On disait qu'il ressemblait à Byron - par la tête, car il était irréprochable quant aux pieds - , mais un Byron à moustaches et à favoris, un Byron impassible,
  qui aurait vécu mille ans sans vieillir.
6 Anglais, à coup sûr, Phileas Fogg n'était peut-être pas Londonner. On ne l'avait jamais vu ni à la Bourse, ni à la Banque, ni dans aucun des comptoirs de la
  Cité. Ni les bassins ni les docks de Londres n'avaient jamais reçu un navire ayant pour armateur Phileas Fogg. Ce gentleman ne figurait dans aucun comité
  d'administration. Son nom n'avait jamais retenti dans un collège d'avocats, ni au Temple, ni à Lincoln's-inn, ni à Gray's-inn. Jamais il ne plaida ni à la Cour
  du chancelier, ni au Banc de la Reine, ni à l'Échiquier, ni en Cour ecclésiastique. Il n'était ni industriel, ni négociant, ni marchand, ni agriculteur. Il ne
  faisait partie ni de l'Institution royale de la Grande-Bretagne, ni de l'Institution de Londres, ni de l'Institution des Artisans, ni de l'Institution Russell,
  ni de l'Institution littéraire de l'Ouest, ni de l'Institution du Droit, ni de cette Institution des Arts et des Sciences réunis, qui est placée sous le
  patronage direct de Sa Gracieuse Majesté. Il n'appartenait enfin à aucune des nombreuses sociétés qui pullulent dans la capitale de l'Angleterre, depuis la
  Société de l'Armonica jusqu'à la Société entomologique, fondée principalement dans le but de détruire les insectes nuisibles.
7 Phileas Fogg était membre du Reform-Club, et voilà tout.
8 A qui s'étonnerait de ce qu'un gentleman aussi mystérieux comptât parmi les membres de cette honorable association, on répondra qu'il passa sur la
  recommandation de MM. Baring frères, chez lesquels il avait un crédit ouvert. De là une certaine « surface », due à ce que ses chèques étaient régulièrement
  payés à vue par le débit de son compte courant invariablement créditeur.
9 Ce Phileas Fogg était-il riche ? Incontestablement. Mais comment il avait fait fortune, c'est ce que les mieux informés ne pouvaient dire, et Mr. Fogg était le
  dernier auquel il convint de s'adresser pour l'apprendre. En tout cas, il n'était prodigue de rien, mais non avare, car partout où il manquait un appoint pour
  une chose noble, utile ou généreuse, il l'apportait silencieusement et même anonymement.
10 En somme, rien de moins communicatif que ce gentleman. Il parlait aussi peu que possible, et semblait d'autant plus mystérieux qu'il était silencieux. Cependant
  sa vie était à jour, mais ce qu'il faisait était si mathématiquement toujours la même chose, que l'imagination, mécontente, cherchait au-delà.
11 Avait-il voyagé ? C'était probable, car personne ne possédait mieux que lui la carte du monde. Il n'était endroit si reculé dont il ne parût avoir une
  connaissance spéciale. Quelquefois, mais en peu de mots, brefs et clairs, il redressait les mille propos qui circulaient dans le club au sujet des voyageurs
  perdus ou égarés ; il indiquait les vraies probabilités, et ses paroles s'étaient trouvées souvent comme inspirées par une seconde vue, tant l'événement
  finissait toujours par les justifier. C'était un homme qui avait dû voyager partout, - en esprit, tout au moins.
12 Ce qui était certain toutefois, c'est que, depuis de longues années, Phileas Fogg n'avait pas quitté Londres. Ceux qui avaient l'honneur de le connaître un peu
  plus que les autres attestaient que - si ce n'est sur ce chemin direct qu'il parcourait chaque jour pour venir de sa maison au club - personne ne pouvait
  prétendre l'avoir jamais vu ailleurs. Son seul passe-temps était de lire les journaux et de jouer au whist. A ce jeu du silence, si bien approprié à sa nature,
  il gagnait souvent, mais ses gains n'entraient jamais dans sa bourse et figuraient pour une somme importante à son budget de charité. D'ailleurs, il faut le
```

Normal text file      length: 421877 lines: 2061      Ln: 1 Col: 1 Sel: 0/0      Windows (CR LF)      **ANSI**      INS

# Conversion de l'encodage d'un texte



# Conversion de l'encodage du texte

```
80jours.txt x
1 Chapitre I
2 DANS LEQUEL PHILEAS FOGG ET PASSEPARTOUT S'ACCEPTENT RÉCIPROQUEMENT L'UN COMME MAÎTRE, L'AUTRE COMME DOMESTIQUE
3 En l'année 1872, la maison portant le numéro 7 de Saville-row, Burlington Gardens - maison dans laquelle Sheridan mourut en 1814 - , était habitée par Phileas
  Fogg, esq., l'un des membres les plus singuliers et les plus remarqués du Reform-Club de Londres, bien qu'il semblât prendre à tâche de ne rien faire qui pût
  attirer l'attention.
4 A l'un des plus grands orateurs qui honorent l'Angleterre, succédait donc ce Phileas Fogg, personnage énigmatique, dont on ne savait rien, sinon que c'était un
  fort galant homme et l'un des plus beaux gentlemen de la haute société anglaise.
5 On disait qu'il ressemblait à Byron - par la tête, car il était irréprochable quant aux pieds - , mais un Byron à moustaches et à favoris, un Byron impassible,
  qui aurait vécu mille ans sans vieillir.
6 Anglais, à coup sûr, Phileas Fogg n'était peut-être pas Londonner. On ne l'avait jamais vu ni à la Bourse, ni à la Banque, ni dans aucun des comptoirs de la
  Cité. Ni les bassins ni les docks de Londres n'avaient jamais reçu un navire ayant pour armateur Phileas Fogg. Ce gentleman ne figurait dans aucun comité
  d'administration. Son nom n'avait jamais retenti dans un collège d'avocats, ni au Temple, ni à Lincoln's-inn, ni à Gray's-inn. Jamais il ne plaida ni à la Cour
  du chancelier, ni au Banc de la Reine, ni à l'Échiquier, ni en Cour ecclésiastique. Il n'était ni industriel, ni négociant, ni marchand, ni agriculteur. Il ne
  faisait partie ni de l'Institution royale de la Grande-Bretagne, ni de l'Institution de Londres, ni de l'Institution des Artisans, ni de l'Institution Russell,
  ni de l'Institution littéraire de l'Ouest, ni de l'Institution du Droit, ni de cette Institution des Arts et des Sciences réunis, qui est placée sous le
  patronage direct de Sa Gracieuse Majesté. Il n'appartenait enfin à aucune des nombreuses sociétés qui pullulent dans la capitale de l'Angleterre, depuis la
  Société de l'Armonica jusqu'à la Société entomologique, fondée principalement dans le but de détruire les insectes nuisibles.
7 Phileas Fogg était membre du Reform-Club, et voilà tout.
8 À qui s'étonnerait de ce qu'un gentleman aussi mystérieux comptât parmi les membres de cette honorable association, on répondra qu'il passa sur la
  recommandation de MM. Baring frères, chez lesquels il avait un crédit ouvert. De là une certaine « surface », due à ce que ses chèques étaient régulièrement
  payés à vue par le débit de son compte courant invariablement créditeur.
9 Ce Phileas Fogg était-il riche ? Incontestablement. Mais comment il avait fait fortune, c'est ce que les mieux informés ne pouvaient dire, et Mr. Fogg était le
  dernier auquel il convint de s'adresser pour l'apprendre. En tout cas, il n'était prodigue de rien, mais non avare, car partout où il manquait un appoint pour
  une chose noble, utile ou généreuse, il l'apportait silencieusement et même anonymement.
10 En somme, rien de moins communicatif que ce gentleman. Il parlait aussi peu que possible, et semblait d'autant plus mystérieux qu'il était silencieux. Cependant
  sa vie était à jour, mais ce qu'il faisait était si mathématiquement toujours la même chose, que l'imagination, mécontente, cherchait au-delà.
11 Avait-il voyagé ? C'était probable, car personne ne possédait mieux que lui la carte du monde. Il n'était endroit si reculé dont il ne parût avoir une
  connaissance spéciale. Quelquefois, mais en peu de mots, brefs et clairs, il redressait les mille propos qui circulaient dans le club au sujet des voyageurs
  perdus ou égarés ; il indiquait les vraies probabilités, et ses paroles s'étaient trouvées souvent comme inspirées par une seconde vue, tant l'événement
  finissait toujours par les justifier. C'était un homme qui avait dû voyager partout, - en esprit, tout au moins.
12 Ce qui était certain toutefois, c'est que, depuis de longues années, Phileas Fogg n'avait pas quitté Londres. Ceux qui avaient l'honneur de le connaître un peu
  plus que les autres attestaient que - si ce n'est sur ce chemin direct qu'il parcourait chaque jour pour venir de sa maison au club - personne ne pouvait
  prétendre l'avoir jamais vu ailleurs. Son seul passe-temps était de lire les journaux et de jouer au whist. A ce jeu du silence, si bien approprié à sa nature,
  il gagnait souvent, mais ses gains n'entraient jamais dans sa bourse et figuraient pour une somme importante à son budget de charité. D'ailleurs, il faut le
```

Normal text file      length: 432 513    lines: 2061    Ln: 1   Col: 1   Sel: 0 | 0    Windows (CR LF)    **UCS-2 LE BOM**    INS

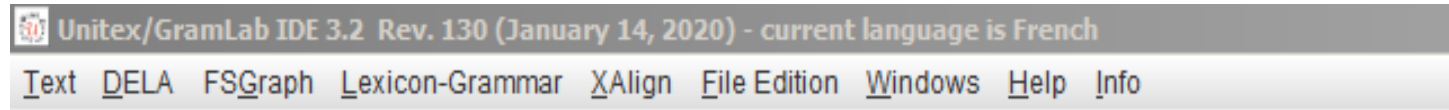
**Le texte a été converti en UTF-16LE BOM -  
encodage par défaut.**

# Analyse avec Winhex - Conversion Windows-1252 (ANSI) → UTF-16LE BOM

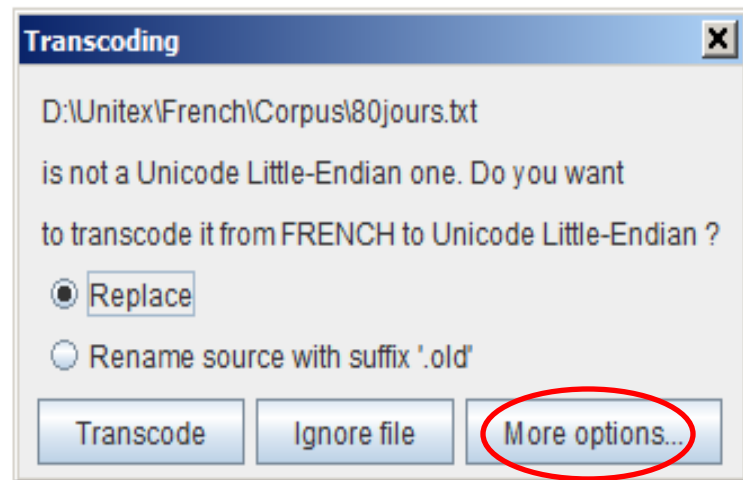
0	1	2	3	4	5	6	7	8	9	A	B	
FF	FE	43	00	68	00	61	00	70	00	69	00	ÿþC.h.a.p.i.
74	00	72	00	65	00	20	00	49	00	20	00	t.r.e. .I. .
0D	00	0A	00	44	00	41	00	4E	00	53	00	. . . .D.A.N.S.
20	00	4C	00	45	00	51	00	55	00	45	00	. .L.E.Q.U.E.
4C	00	20	00	50	00	48	00	49	00	4C	00	L. .P.H.I.L.
45	00	41	00	53	00	09	00	09	00	09	00	E.A.S. . . . .
09	00	09	00	09	00	20	00	46	00	4F	00	. . . . .F.O.
47	00	47	00	20	00	45	00	54	00	20	00	G.G. .E.T. .
50	00	41	00	53	00	53	00	45	00	50	00	P.A.S.S.E.P.
41	00	52	00	54	00	4F	00	55	00	54	00	A.R.T.O.U.T.
20	00	53	00	27	00	41	00	43	00	43	00	.S.' .A.C.C.
45	00	50	00	54	00	45	00	4E	00	54	00	E.P.T.E.N.T.
20	00	52	00	C9	00	43	00	49	00	50	00	.R.É.C.I.P.
52	00	4F	00	51	00	55	00	45	00	4D	00	R.O.Q.U.E.M.
45	00	4E	00	54	00	20	00	4C	00	27	00	E.N.T. .L.'
55	00	4E	00	20	00	43	00	4F	00	4D	00	U.N. .C.O.M.
4D	00	45	00	20	00	4D	00	41	00	CE	00	M.E. .M.A.Î.
54	00	52	00	45	00	2C	00	20	00	4C	00	T.R.E. . .L.
27	00	41	00	55	00	54	00	52	00	45	00	' .A.U.T.R.E.
20	00	43	00	4F	00	4D	00	4D	00	45	00	.C.O.M.M.E.
20	00	44	00	4F	00	4D	00	45	00	53	00	.D.O.M.E.S.
54	00	49	00	51	00	55	00	45	00	20	00	T.I.Q.U.E. .



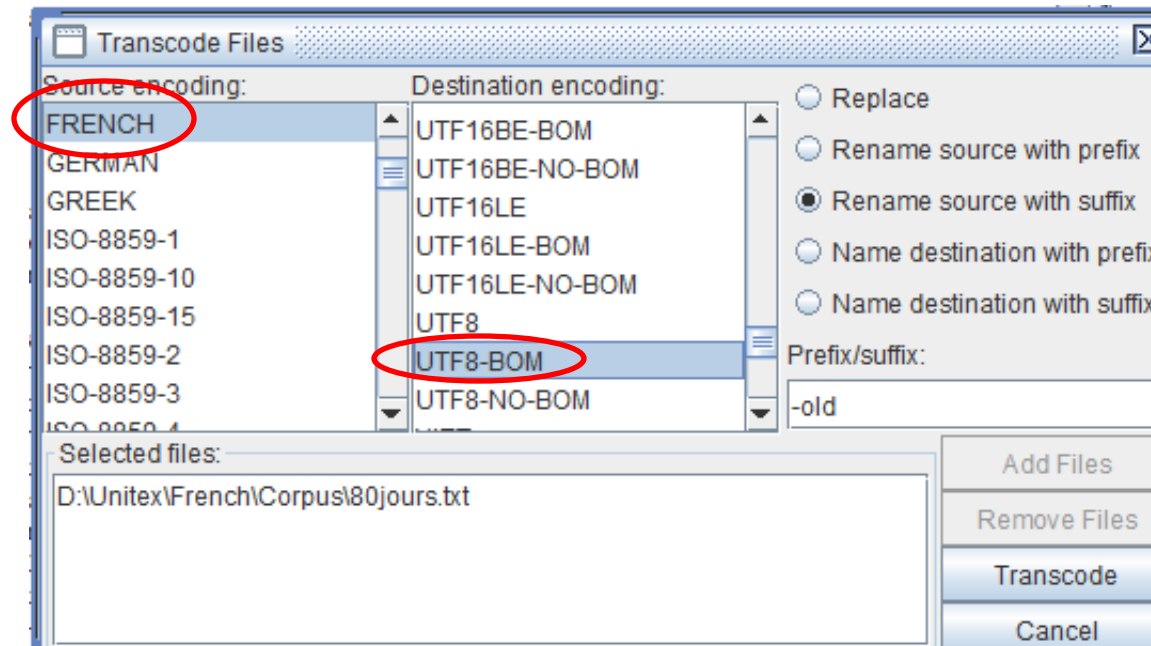
# Conversion de l'encodage d'un texte



## Repartir de l'ANSI



# Conversion de l'encodage d'un texte



# Analyse avec Winhex - Conversion Windows-1252 (ANSI) → UTF-8 BOM

0	1	2	3	4	5	6	7	8	9	A	B	
EF	BB	BF	43	68	61	70	69	74	72	65	20	File >> Chapitre
49	28	0D	0A	44	41	4E	53	20	4C	45	51	1. . . DANS LEQ
55	45	4C	20	50	48	49	4C	45	41	53	20	UEL PHILEAS
46	4F	47	47	20	45	54	20	50	41	53	53	FOGG ET PASS
45	50	41	52	54	4F	55	54	20	53	27	41	EPARTOUT S'A
43	43	45	50	54	45	4E	54	20	52	C3	89	CCEPTENT RÃ
43	49	50	52	4F	51	55	45	4D	45	4E	54	CIPROQUEMENT
20	4C	27	55	4E	20	43	4F	4D	4D	45	20	L'UN COMME
4D	41	C3	8E	54	52	45	2C	20	4C	27	41	MAÃTRE, L'A
55	54	52	45	20	43	4F	4D	4D	45	20	44	UTRE COMME D
4F	4D	45	53	54	49	51	55	45	20	0D	0A	OMESTIQUE ..
45	6E	20	6C	27	61	6E	6E	C3	A9	65	20	En l'annÃe
31	38	37	32	2C	20	6C	61	20	6D	61	69	1872, la mai
73	6F	6E	20	70	6F	72	74	61	6E	74	20	son portant
6C	65	20	6E	75	6D	C3	A9	72	6F	20	37	le numÃro 7
20	64	65	20	53	61	76	69	6C	6C	65	2D	de Saville-
72	6F	77	2C	20	42	75	72	6C	69	6E	67	row, Burling
74	6F	6E	20	47	61	72	64	65	6E	73	20	ton Gardens
2D	20	6D	61	69	73	6F	6E	20	64	61	6E	- maison dan
73	20	6C	61	71	75	65	6C	6C	65	20	53	s laquelle S
68	65	72	69	64	61	6E	20	6D	6F	75	72	heridan mour
75	74	20	65	6E	20	31	38	31	34	20	2D	ut en 1814 -
20	2C	20	C3	A9	74	61	69	74	20	68	61	, Ãtait ha
62	69	74	C3	A9	65	20	70	61	72	20	50	bitÃe par P
68	69	6C	65	61	73	20	46	6F	67	67	2C	hileas Fogg,
20	65	73	71	2E	2C	20	6C	27	75	6E	20	esq., l'un
64	65	73	20	6D	65	6D	62	72	65	73	20	des membres
6C	65	73	20	70	6C	75	73	20	73	69	6E	les plus sin
67	75	6C	69	65	72	73	20	65	74	20	6C	guliers et l
65	73	20	70	6C	75	73	20	72	65	6D	61	es plus rema
72	71	75	C3	A9	73	20	64	75	20	52	65	rquÃs du Re
66	6F	72	6D	2D	43	6C	75	62	20	64	65	form-Club de
20	4C	6F	6E	64	72	65	73	2C	20	62	69	Londres, bi
65	6E	20	71	75	27	69	6C	20	73	65	6D	en qu'il sem
62	6C	C3	A2	74	20	70	72	65	6E	64	72	blÃet prendr
65	20	C3	A0	20	74	C3	A2	63	68	65	20	e Ã tÃche

# Exercice 1

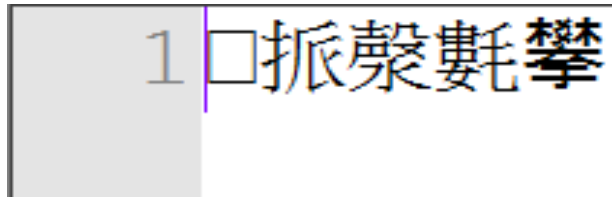
- Quel est le BOM des suites hexadécimales suivantes ?
- Ya-t-il des anomalies dans ces suites hexadécimales ? Si oui, expliquez-les.
- FE FF 00 E9 00 63 00 6F 00 6C 00 65
- FF FE E9 00 63 00 6F 00 6C 00 65 00
- EF BB BF 3C 00 21 00 44 00 4F 00 43 00 54 00
- FE FF 3C 00 21 00 44 00 4F 00 43 00 54 00
- FE EE 00 E9 00 63 00 6F 00 6C 00 65
- FF FE 00 E9 00 63 00 6F 00 6C 00 65
- EF BB BF C3 A9 63 6F 6C 65 20 73 75 70 C3 A9 72 69  
65 75 72 65
- EF BB BF C3 A9 00 6C C3 A8 00 76 00 65

## Exercice 2

- Créer un fichier encodé en "UTF-8" sans BOM (Taille = 6 octets), contenant le mot **école**
- Écrire un script Python permettant de :
- 1- lire ce fichier et générer un fichier encodé en "UTF-16 Little-Endian" avec BOM (Taille = 12 octets), contenant ce mot.
- 2- lire ce fichier et générer un fichier encodé en "UTF-16 Big-Endian" avec BOM (Taille = 12 octets), contenant ce mot.

## Exercice 3

- Créer un fichier encodé en "UTF-8" sans BOM (Taille = 6 octets), contenant le mot **école**
- Écrire un script Python permettant de générer ce résultat.
- Comment expliquer le résultat ci-dessous obtenu lors d'une conversion de l'encodage source vers un encodage UTF-16 ?



1 □ 校 園 學 校

## Exercice 4

- Créer un fichier encodé en "ANSI" (Taille = 5 octets), contenant le mot **école**
- Écrire un script Python permettant de :
- - lire ce fichier et générer un fichier encodé en "UTF-8" avec BOM (Taille = 9 octets), contenant ce mot.