

- Université de Bouira

Codage et représentation de l'information

MI, semestre 1

Taha Zerrouki

Programme

- **Représentation des nombres**
 - **BCD**
 - **Exces 3**
 - **Code Gray**
- **Représentation des caractères**
 - **ASCII**
 - **Unicode**

Codage des caractères
ترميز الحروف

Codage d'information

A

65

01000001

Codage d'information

Ada b

A	d	a		b
---	---	---	--	---

65	100	97	32	98
----	-----	----	----	----

01000001	01100100	01100001	00100000	01100010
----------	----------	----------	----------	----------

Codage des caractères

Un nombre correspond à une Lettre

65 <====> A

66 <====> B

Codage des caractères

Un nombre correspond à une Lettre

65 <====> A

66 <====> B

Code ASCII

American Standard Code for Information
Interchange

الترميز المعياري الأمريكي لتبادل المعلومات

ASCII

- 7 bits => 128 caractères
- + 1 bit de vérification

ASCII

!"#\$%&'()*+,-./0123456789:;<=>?

@ABCDEFGHIJKLMNOPQRSTUVWXYZ[

\]^_

`abcdefghijklmnopqrstuvwxyz{|}

codes

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

Caractères spéciaux

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0x	NUL	SOH	STX	ETX	EOT	ENQ	ACK	<u>BEL</u>	<u>BS</u>	<u>HT</u>	<u>LF</u>	VT	FF	<u>CR</u>	SO	SI
1x	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	<u>ESC</u>	FS	GS	RS	US
2x	<u>SP</u>	<u>!</u>	<u>"</u>	<u>#</u>	<u>\$</u>	<u>%</u>	<u>&</u>	<u>'</u>	<u>(</u>	<u>)</u>	<u>*</u>	<u>+</u>	<u>,</u>	<u>-</u>	<u>.</u>	<u>/</u>

Caractères spéciaux

- SP : 20 espace
- CR : 13 retour à la ligne
- HT : 09 Tabulation
- BS ; 08 Back space
- BEL : 07 sonnette
- NUL : 00 null
- DEL : 127 supprimer
-

Space

Enter

Tab

Backspace

Exercice

- Décoder le message suivant
- 42 6f 75 69 72 61 20 31 30

Exercice

- Décoder le message suivant
- 42 6f 75 69 72 61 20 31 30
- B o u i R a 1 0

Exercice

- Coder le message suivant
- L'Algérie

Exercice

- Coder le message suivant
- L ' A l g é r i e
- 4c 27 41 6c 67 ?? 72 69 65

Problèmes

- Il Représente que l'anglais
 - Manque des caractères accentués
 - Pas des caractères multilingues
- يمثل الإنجليزية فقط
 - لا حروف منقطة
 - لا حروف من لغات غير لاتينية

ASCII étendu














- Utiliser 8 bits
- 8 bits = 256 caractères
- 0-127 => ascii
- 128-255 => extension توسيع

ASCII étendu

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
20		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
60	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
80			,	f	„	...	†	‡	^	‰	Š	<	Œ			
90		‘	’	“	”	•	—	—	~	™	š	>	œ			ÿ
A0		¡	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	­	®	¯
B0	°	±	²	³	´	µ	¶	·	,	¹	º	»	¼	½	¾	¿
C0	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D0	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E0	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F0	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

ASCII étendu Arabe

708 MS-DOS ARABIC ASMO

	20	30	40	50	60	70	80	90	A0	B0	C0	D0	E0	F0
0	 32	0 48	@ 64	P 80	` 96	p 112	 128	 144	 NOT USED 160	 176	# 192	ذ 208	— 224	 240
1	! 33	1 49	A 65	Q 81	a 97	q 113	† 129	‡ 145	⊥ 161	 177	ء 193	ر 209	ف 225	’ 241
2	" 34	2 50	B 66	R 82	b 98	r 114	é 130	ك 146	⌈ 162	 178	آ 194	ز 210	ق 226	° 242
3	# 35	3 51	C 67	S 83	c 99	s 115	â 131	ô 147	⌋ 163	‡ 179	أ 195	س 211	ك 227	 NOT USED 243
4	\$ 36	4 52	D 68	T 84	d 100	t 116	‡ 132	‡ 148	⌌ 164	= 180	ؤ 196	ش 212	ل 228	 NOT USED 244
5	% 37	5 53	E 69	U 85	e 101	u 117	à 133	‡ 149	— 165	‡ 181	إ 197	ص 213	م 229	 NOT USED 245
6	& 38	6 54	F 70	V 86	f 102	v 118	‡ 134	û 150	⊕ 166	⊖ 182	ئ 198	ض 214	ن 230	 NOT USED 246
7	' 39	7 55	G 71	W 87	g 103	w 119	ç 135	ù 151	‡ 167	‡ 183	ا 199	ط 215	ه 231	 NOT USED 247
8	(40	8 56	H 72	X 88	h 104	x 120	ê 136	‡ 152	‡ 168	‡ 184	ب 200	ظ 216	و 232	 NOT USED 248
9) 41	9 57	I 73	Y 89	i 105	y 121	ë 137	‡ 153	‡ 169	‡ 185	ة 201	ع 217	ي 233	‡ 249
A	* 42	: 58	J 74	Z 90	j 106	z 122	è 138	 NOT USED 154	‡ 170	‡ 186	ت 202	غ 218	ي 234	‡ 250

ASCII étendu russe

—		┐	┌	└	┘	┌	└	┐	┘	┐	■	■	■	■	■
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
▤	▥	▦	┐	■	●	√	≈	≤	≥	nbsp	┘	◦	²	•	÷
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
=		ƒ	ë	π	⌌	ƒ	π	π	⌌	⌌	⌌	⌌	⌌	⌌	⌌
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
	⌌	ƒ	Ë		⌌	⌌	π	π	⌌	⌌	⌌	⌌	⌌	⌌	©
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
ю	а	б	ц	д	е	ф	г	х	и	й	к	л	м	н	о
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
п	я	р	с	т	у	ж	в	ь	ы	з	ш	э	щ	ч	ъ
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
Ю	А	Б	Ц	Д	Е	Ф	Г	Х	И	Й	К	Л	М	Н	О
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
П	Я	Р	С	Т	У	Ж	В	Ь	Ы	З	Ш	Э	Щ	Ч	Ъ
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

ASCII étendu thaïlandais

6-	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
	0060	0061	0062	0063	0064	0065	0066	0067	0068	0069	006A	006B	006C	006D	006E	006F
7-	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
	0070	0071	0072	0073	0074	0075	0076	0077	0078	0079	007A	007B	007C	007D	007E	007F
8-																
9-																
A-	'	ก	ข	ฃ	ค	ฅ	ฉ	ง	จ	ฉ	ช	ซ	ฌ	ญ	ฎ	ฏ
	0E48	0E01	0E02	0E03	0E04	0E05	0E06	0E07	0E08	0E09	0E0A	0E0B	0E0C	0E0D	0E0E	0E0F
B-	ฐ	ฑ	ฒ	ณ	ด	ต	ถ	ท	ธ	น	บ	ป	ผ	ฝ	พ	ฟ
	0E10	0E11	0E12	0E13	0E14	0E15	0E16	0E17	0E18	0E19	0E1A	0E1B	0E1C	0E1D	0E1E	0E1F
C-	ภ	ม	ย	ร	ล	ฬ	ว	ศ	ษ	ส	ห	ฬ	อ	ฮ	า	ะ
	0E20	0E21	0E22	0E23	0E24	0E25	0E26	0E27	0E28	0E29	0E2A	0E2B	0E2C	0E2D	0E2E	0E2F
D-	ะ	ั	า	ำ	อ	ะ	ั	า	ำ	อ	ะ	ั	า	ำ	อ	ะ
	0E30	0E31	0E32	0E33	0E34	0E35	0E36	0E37	0E38	0E39	0E3A	0E40	0E4A	0E4B	0E4C	0E3F
E-	เ	แ	โ	ใ	เ	แ	โ	ใ	เ	แ	โ	ใ	เ	แ	โ	ใ
	0E40	0E41	0E42	0E43	0E44	0E45	0E46	0E47	0E48	0E49	0E4A	0E4B	0E4C	0E4D	0E4E	0E4F
F-	อ	ะ	ั	า	ำ	อ	ะ	ั	า	ำ	อ	ะ	ั	า	ำ	อ
	0E50	0E51	0E52	0E53	0E54	0E55	0E56	0E57	0E58	0E59	0E5A	0E5B	00A2	00AC	00A6	00A0

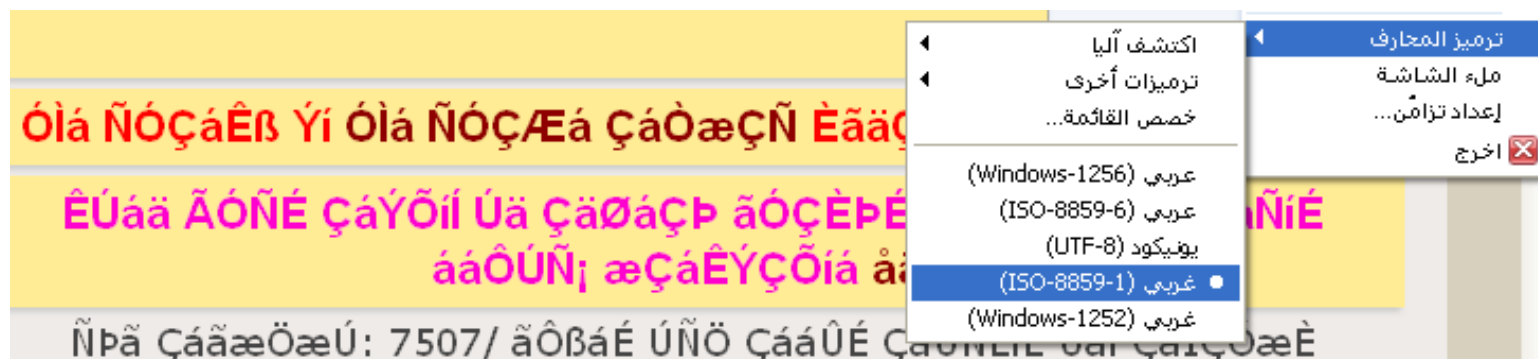
Problèmes

- Multitude des codes تعدد الترميزات
- Langues avec des nombreux caractères
- مشكلة اللغات ذات الحروف الكثيرة

Problèmes

- Multitude des codes تعدد الترميزات
- Langues avec des nombreux caractères
- مشكلة اللغات ذات الحروف الكثيرة

Différents codes



Différents codes



الترميز العالمي الموحد

Unicode

Unicode

16 bits = 65 535 caractères

Multilingue

Unicode

A

65

ض

1560

𐎠

11619

鴉

40201

Unicode

a

LATIN SMALL
LETTER A

0061

ã

LATIN SMALL
LETTER A
WITH TILDE

00E3

ي

ARABIC
LETTER E

06D0

不

CJK UNIFIED
IDEOGRAPH-
4E0D

4E0D



SNOWMAN

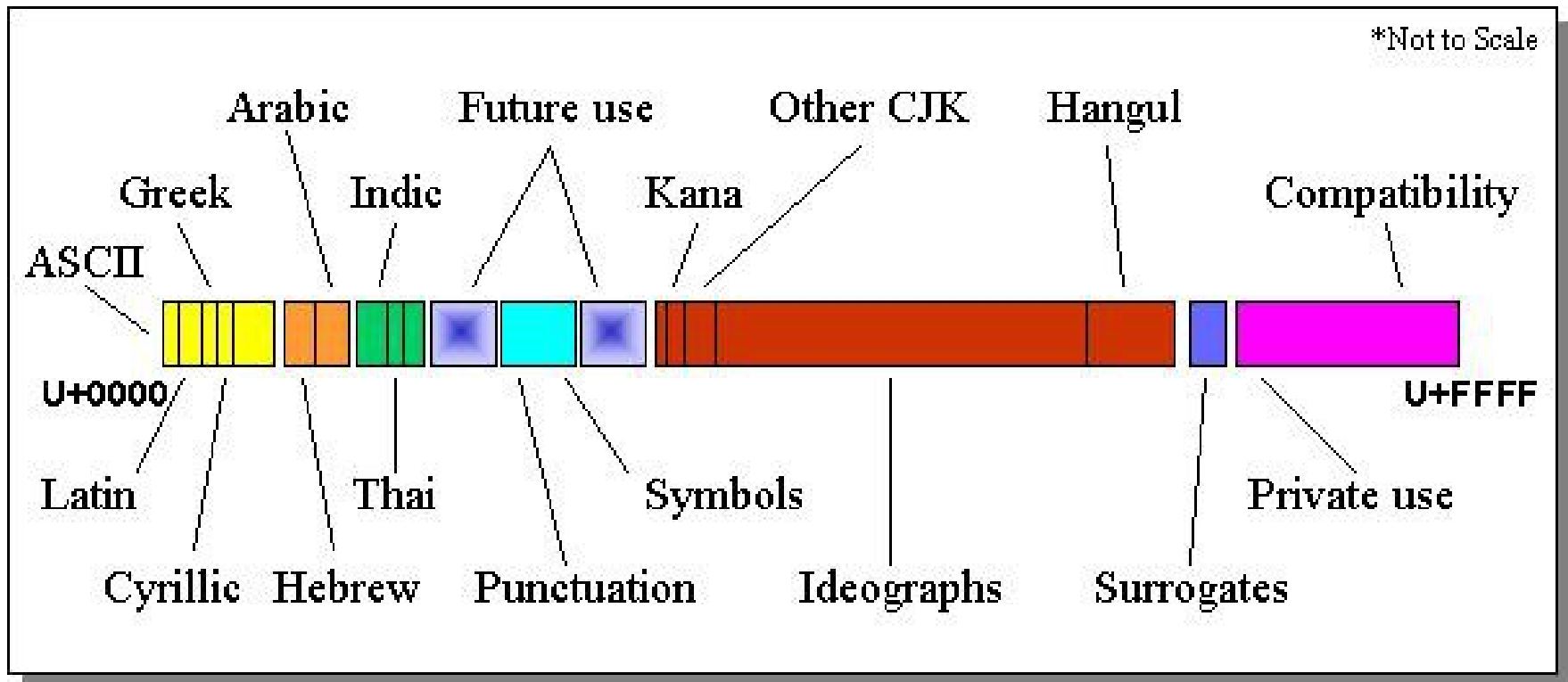
2603

Unicode

- Codage multilingue ترميز متعدد اللغات
- 2 octets = 16 bits
- Extensible قابل للإثراء

Unicode.org

Unicode

































Unicode chinois

03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	UTF8
局	处	夆	夆	备	条	变	夂	夂	夂	复	爰	夏	E5A480
<u>5903</u>	<u>5904</u>	<u>5905</u>	<u>5906</u>	<u>5907</u>	<u>5908</u>	<u>5909</u>	<u>590A</u>	<u>590B</u>	<u>590C</u>	<u>590D</u>	<u>590E</u>	<u>590F</u>	
璽	夔	夕	外	夙	夙	夙	多	彗	夜	姓	訖	够	E5A490
<u>5913</u>	<u>5914</u>	<u>5915</u>	<u>5916</u>	<u>5917</u>	<u>5918</u>	<u>5919</u>	<u>591A</u>	<u>591B</u>	<u>591C</u>	<u>591D</u>	<u>591E</u>	<u>591F</u>	
𡗗	貴	夥	夥	大	夊	天	太	夫	夊	夭	央	夊	E5A4A0
<u>5923</u>	<u>5924</u>	<u>5925</u>	<u>5926</u>	<u>5927</u>	<u>5928</u>	<u>5929</u>	<u>592A</u>	<u>592B</u>	<u>592C</u>	<u>592D</u>	<u>592E</u>	<u>592F</u>	
𡗗	头	夊	夊	夷	夸	夹	夺	夺	夊	夊	夊	夊	E5A4B0
<u>5933</u>	<u>5934</u>	<u>5935</u>	<u>5936</u>	<u>5937</u>	<u>5938</u>	<u>5939</u>	<u>593A</u>	<u>593B</u>	<u>593C</u>	<u>593D</u>	<u>593E</u>	<u>593F</u>	

Tifinagh














































































The Unicode Standard
Version 6.3

	2D3	2D4	2D5	2D6	2D7
0	 2D30	 2D40	 2D50	 2D60	 2D70
1	 2D31	 2D41	 2D51	 2D61	
2	 2D32	 2D42	 2D52	 2D62	
3	 2D33	 2D43	 2D53	 2D63	
4	 2D34	 2D44	 2D54	 2D64	
5	 2D35	 2D45	 2D55	 2D65	








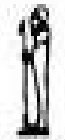



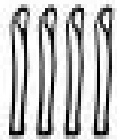























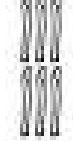











Unicode Math

2200	\forall	\mathbb{C}	∂	\exists	\nexists	\emptyset	Δ	∇	\in	\notin	ϵ
2210	\prod	\sum	$-$	\mp	\pm	$/$	\setminus	$*$	\circ	\bullet	$\sqrt{}$
2220	\angle	\sphericalangle	\triangleleft	$ $	\vdash	\parallel	\nparallel	\wedge	\vee	\cap	\cup
2230	\oint	<u>2231</u>	\oint	\oint	\therefore	\therefore	\therefore	\therefore	$\dot{+}$	$\dot{-}$	$\dot{=}$
2240	\wr	\nwr	\approx	\approx	\napprox	\approx	\napprox	\napprox	\approx	\napprox	\approx
2250	\doteq	\doteq	$\dot{=}$	$\dot{=}$	$\dot{=}$	$\dot{=}$	$\dot{=}$	$\dot{=}$	$\dot{=}$	$\dot{=}$	$\dot{=}$
2260	\neq	\equiv	\neq	\equiv	\neq	\equiv	\neq	\equiv	\neq	\equiv	\neq
2270	\nless	\nless	\lessgtr	\lessgtr	\lessgtr	\lessgtr	\lessgtr	\lessgtr	\lessgtr	\lessgtr	\lessgtr
2280	\nless	\nless	\subset	\subset	$\not\subset$	$\not\subset$	\subset	\subset	$\not\subset$	$\not\subset$	\subset
2290	\sqsubset	\sqsubset	\sqsubset	\sqsubset	\sqsubset	\oplus	\oplus	\otimes	\oslash	\odot	\odot

Unicode symbols

	0	1	2	3	4	5	6	7	8	9	A
U+260x											
U+261x											
U+262x											
U+263x											
U+264x											
U+265x											
U+266x											

Unicode Hiéroglyphique

1300	1301	1302	1303	1304	1305	1306	1307	1308	1309	130A	130B	1
 13000	 13010	 13020	 13030	 13040	 13050	 13060	 13070	 13080	 13090	 130A0	 130B0	
 13001	 13011	 13021	 13031	 13041	 13051	 13061	 13071	 13081	 13091	 130A1	 130B1	
 13002	 13012	 13022	 13032	 13042	 13052	 13062	 13072	 13082	 13092	 130A2	 130B2	
 13003	 13013	 13023	 13033	 13043	 13053	 13063	 13073		 13093		 130B3	 13004

L'arabe et l'Unicode

العربية الأساسية Arabe standard

العلامات القرآنية Coranique Marques

العربية الموسعة Arabe étendue

- الفارسية
- الأردو
- البشتو
- الكردية
- اللغات الإفريقية

التراكيب مثل لا، لم Ligature

Unicode arabe

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F
0	 0600	ص 0610	ي 0620	ذ 0630	- 0640	◌ِ 0650	◌ْ 0660	◌ُ 0670	پ 0680	ت 0690	ث 06A0	گ 06B0	ه 06C0	ي 06D0	◌ُ 06E0	◌ْ 06F0
1	 0601	ع 0611	ء 0621	ر 0631	ف 0641	◌ِ 0651	◌ْ 0661	أ 0671	خ 0681	ر 0691	ف 06A1	گ 06B1	ه 06C1	ي 06D1	◌ُ 06E1	◌ْ 06F1
2	 0602	◌ِ 0612	آ 0622	ز 0632	ق 0642	◌ُ 0652	◌ْ 0662	أ 0672	خ 0682	ر 0692	ب 06A2	گ 06B2	ه 06C2	ے 06D2	◌ُ 06E2	◌ْ 06F2
3	 0603	ض 0613	أ 0623	س 0633	ك 0643	◌ِ 0653	◌ْ 0663	إ 0673	ج 0683	ر 0693	ف 06A3	گ 06B3	ت 06C3	ے 06D3	◌ُ 06E3	◌ْ 06F3
4	 0604	◌ِ 0614	ؤ 0624	ش 0634	ل 0644	◌ُ 0654	◌ْ 0664	ء 0674	ج 0684	ر 0694	ف 06A4	گ 06B4	و 06C4	- 06D4	◌ُ 06E4	◌ْ 06F4
5		ط 0615	إ 0625	ص 0635	م 0645	◌ِ 0655	◌ْ 0665	ه 0675	أ 0685	ر 0695	پ 06A5	ل 06B5	و 06C5	ه 06D5	◌ُ 06E5	◌ْ 06F5
6	٣ 0606	ال 0616	ئ 0626	ض 0636	ن 0646	◌ِ 0656	◌ْ 0666	و 0676	چ 0686	ر 0696	ق 06A6	ن 06B6	و 06C6	ص 06D6	◌ُ 06E6	◌ْ 06F6

Exercice

- Décoder le message suivant
- 062c 0627 0645 0650
0639 0629 064c

Exercice

- Décoder le message suivant
- 062c 0627 0645 0650 0639 0629 064c

62c	627	645	650	639	629	64c
ج	ا	م	كسرة	ع	ة	ضمّتان

Exercice

- Coder le message suivant

البويرة

Exercice

- Coder le message suivant

البويرة

u0627 0644 0628 0648
064a 0652 0631 0629

العلامات القرآنية في اليونيكود

نص قرآني: سورة النمل من آية 25 إلى الآية 27:

أَلَّا يَسْجُدُوا لِلَّهِ الَّذِي يُخْرِجُ الْخَبْءَ فِي السَّمَاوَاتِ وَالْأَرْضِ وَيَعْلَمُ مَا تُخْفُونَ وَمَا
تُعْلِنُونَ ﴿١﴾ اللَّهُ لَا إِلَهَ إِلَّا هُوَ رَبُّ الْعَرْشِ الْعَظِيمِ ﴿٢﴾ قَالَ سَتَنْظُرُونَ أَصَدَقْتُ
أَمْ كُنْتُ مِنَ الْكَاذِبِينَ ﴿٣﴾

Caractères de contrôle

- U+202A: LEFT-TO-RIGHT EMBEDDING
- U+202B: RIGHT-TO-LEFT EMBEDDING
- U+202D: LEFT-TO-RIGHT OVERRIDE
- U+202E: RIGHT-TO-LEFT OVERRIDE
- U+202C: POP DIRECTIONAL FORMATTING
- U+200E: LEFT-TO-RIGHT MARK
- U+200F: RIGHT-TO-LEFT MARK

Affichage

ةيبرعلا

العربية

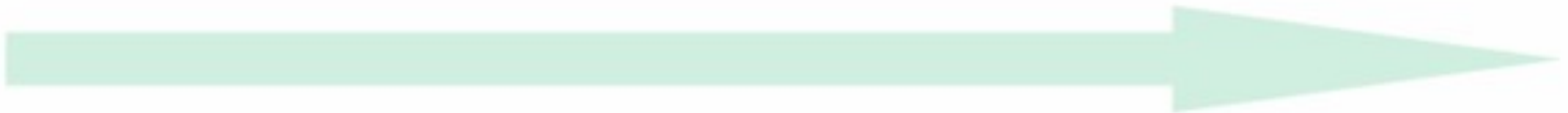
العربية

- ثنائية الاتجاه
- تشبيك الحروف

Bidirectionnalité

الأمير Charles والأميرة Camilla

Camilla أميرة Charles و Charles أمير Camilla



Directions

[illegible]

المنغولية
Mongolien

Directions

- الصينية
- اليابانية
- الكورية

當世界需要溝通時，請用統一碼（Unicode）你現在就應報名將在1997年3月10至12日於德國美姿城（Mainz）召開的第十屆國際統一碼研討會。本次研討會將邀請多位業界專家研討關於全球網際網路及統一碼發展、國際化及本土化、支援統一碼的作業系統及應用程式、字型、文字排版、電腦多國語文化等多項課題。

Caractères de contrôle

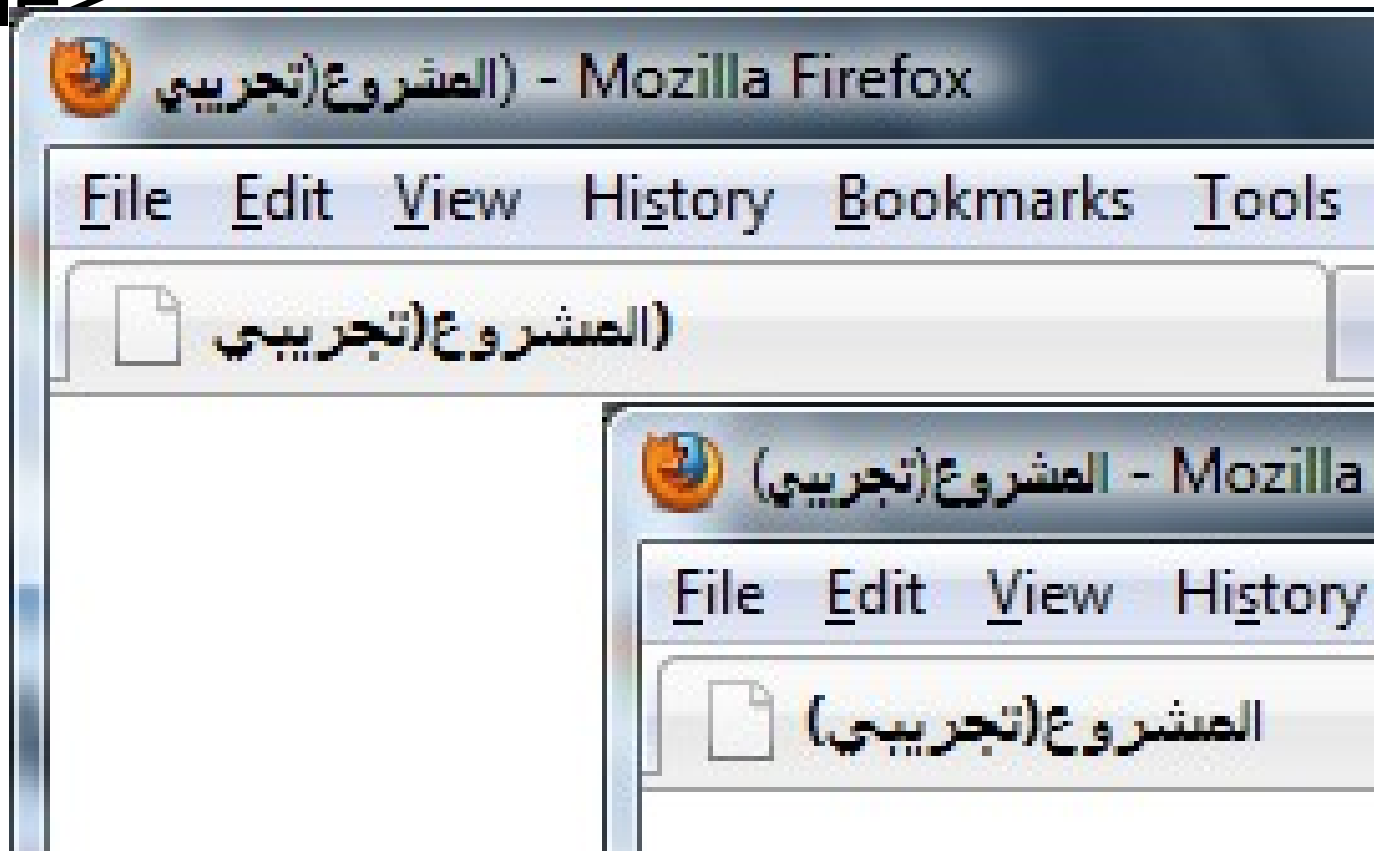
- U+202A: LEFT-TO-RIGHT EMBEDDING
- U+202B: RIGHT-TO-LEFT EMBEDDING
- U+202D: LEFT-TO-RIGHT OVERRIDE
- U+202E: RIGHT-TO-LEFT OVERRIDE
- U+202C: POP DIRECTIONAL FORMATTING
- U+200E: LEFT-TO-RIGHT MARK
- U+200F: RIGHT-TO-LEFT MARK

Example

RLE

PDF

- `<title>‫(تجريبي)‬
</title>`



Affichage

ةيبرعلا

العربية

العربية

- ثنائية الاتجاه

- تشبيك الحروف

Attachement des lettres

Attachement

- اختيار شكل الحرف المناسب حسب السياق
- يجرى عادة على مستوى الخط

لحم

لحم

Hindu

تشابك الحروف في الهندية

ड + ् + क → डक

Représentation BiDi

Dans la mémoire

0	1	2	3	4	5	6
ا	ل	ع	ر	ب	ي	ة

Traitement

Affichage

العربية

Caractères de contrôle d'attachement

- كتابة التاريخ الهجري : 1436 هـ

• هـ = _ + هـ

• هـ = ZWJ + هـ

- Zero width joiner
- **U+200D**

Exemple d'attachement

- ويك لولا الصبر ما كذت ملأت الكيس تبراً

Interdire l'attachement

الجمهورية الجزائرية الديمقراطية الشعبية

ججدش

ج.ج.دش

جZWNJجZWNJدش = < ج.ج.دش

ZWNJ = Zero width non joiner

Représentation

- Unicode utilise 3 représentation :
- **UTF-8**
- Un octet pour l'ascii, 2 à 4 octets pour les autres
- وهو المفضل لدى مبرمجي الويب، حيث يستخدم 1 بايت إذا كانت وتستخدم 2 إلى 4 بايت للرموز، ASCII الرموز موجودة في ترميز المعقدة.
- **UTF-16**
- هذا الترميز يستخدم إما 2 بايت للترميز إذا كانت الرموز موجودة و 4 بايت للرموز الغير (BMP Basic Multilingual Plane) في موجودة.
- **UTF-32** : utilise 4 octets toujours . يستخدم 4 بايت على الدوام

UTF

	A	𐀀	好	𠩺
Code point	U+0041	U+05D0	U+597D	U+233B4
UTF-8	41	D7 90	E5 A5 BD	F0 A3 8E B4
UTF-16	00 41	05 D0	59 7D	D8 4C DF B4
UTF-32	00 00 00 41	00 00 05 D0	00 00 59 7D	00 02 33 B4

UTF

A	Ω	語	卩
00000041	000003A9	00008A9E	00010384

UTF-32

A	Ω	語	卩
0041	03A9	8A9E	D800 DF84

UTF-16

A	Ω	語	卩
41	CE A9	E8 AA 9E	F0 90 8E 84

UTF-8