

In the beginning was
TXT



EBCDIC

BBC010																
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0_	0000	0001	0002	0003	0004	0009	0011	0017	0020	0025	0027	0028	0029	002D	002E	002F
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	0010	0011	0012	0013	0018	0019	001B	001C	001D	001E	001F	001G	001H	001I	001J	001K
1_	0016	0017	0018	0019	001A	001B	001C	001D	001E	001F	001G	001H	001I	001J	001K	001L
	001B	001C	001D	001E	001F	001G	001H	001I	001J	001K	001L	001M	001N	001O	001P	001Q
	001A	001B	001C	001D	001E	001F	001G	001H	001I	001J	001K	001L	001M	001N	001O	001P
2_	0018	0019	001A	001B	001C	001D	001E	001F	001G	001H	001I	001J	001K	001L	001M	001N
	001F	001G	001H	001I	001J	001K	001L	001M	001N	001O	001P	001Q	001R	001S	001T	001U
	001D	001E	001F	001G	001H	001I	001J	001K	001L	001M	001N	001O	001P	001Q	001R	001S
3_	0016	0017	0018	0019	001A	001B	001C	001D	001E	001F	001G	001H	001I	001J	001K	001L
	0014	0015	0016	0017	0018	0019	001A	001B	001C	001D	001E	001F	001G	001H	001I	001J
	001B	001C	001D	001E	001F	001G	001H	001I	001J	001K	001L	001M	001N	001O	001P	001Q
4_	0020	0021	0022	0023	0024	0025	0026	0027	0028	0029	002A	002B	002C	002D	002E	002F
	002B	002C	002D	002E	002F	002A	002B	002C	002D	002E	002F	002A	002B	002C	002D	002E
	002A	002B	002C	002D	002E	002F	002A	002B	002C	002D	002E	002F	002A	002B	002C	002D
5_	0026	0027	0028	0029	002A	002B	002C	002D	002E	002F	002A	002B	002C	002D	002E	002F
	002D	002E	002F	002A	002B	002C	002D	002E	002F	002A	002B	002C	002D	002E	002F	002A
	002B	002C	002D	002E	002F	002A	002B	002C	002D	002E	002F	002A	002B	002C	002D	002E
6_	002D	002E	002F	002A	002B	002C	002D	002E	002F	002A	002B	002C	002D	002E	002F	002A
	002B	002C	002D	002E	002F	002A	002B	002C	002D	002E	002F	002A	002B	002C	002D	002E
	002A	002B	002C	002D	002E	002F	002A	002B	002C	002D	002E	002F	002A	002B	002C	002D
7_	112	113	114	115	116	117	118	119	120	121	0040	0038	0023	0040	0037	0022
	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128
	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
8_	0061	0062	0063	0064	0065	0066	0067	0068	0069	006A	006B	006C	006D	006E	006F	006G
	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144
	006A	006B	006C	006D	006E	006F	006G	006H	006I	006J	006K	006L	006M	006N	006O	006P
9_	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161
	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161
	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160
A_	007E	0073	0074	0075	0076	0077	0078	0079	007A	007B	007C	007D	007E	007F	007G	007H
	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
	007E	0073	0074	0075	0076	0077	0078	0079	007A	007B	007C	007D	007E	007F	007G	007H
B_	005E	177	178	179	180	181	182	183	184	185	005B	005D	005F	005G	005H	005I
	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
	005E	0041	0042	0043	0044	0045	0046	0047	0048	0049	004A	004B	004C	004D	004E	004F
C_	{	A	B	C	D	E	F	G	H	I	SHY					
	007B	0041	0042	0043	0044	0045	0046	0047	0048	0049	004A	004B	004C	004D	004E	004F
	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
D_)	J	K	L	M	N	O	P	Q	R						
	007D	004A	004B	004C	004D	004E	004F	0050	0051	0052	0053	0054	0055	0056	0057	0058
	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
E_	\	S	T	U	V	W	X	Y	Z							
	005C	2007	0053	0054	0055	0056	0057	0058	0059	005A	005B	005C	005D	005E	005F	005G
	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
F_	0	1	2	3	4	5	6	7	8	9						
	0030	0031	0032	0033	0034	0035	0036	0037	0038	0039	0030	0031	0032	0033	0034	0035
	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255
	-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-0	-1	-2	-3	-4	-5

Source: <http://en.wikipedia.org/wiki/EBCDIC>

ASCII





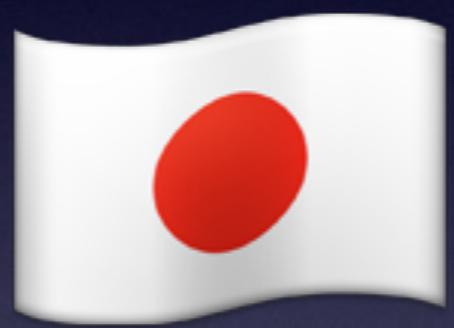
ä, ö, or å and Ø?

Latin-1
ISO/IEC 8859-1

Latin-*

Windows code pages

Then came the €



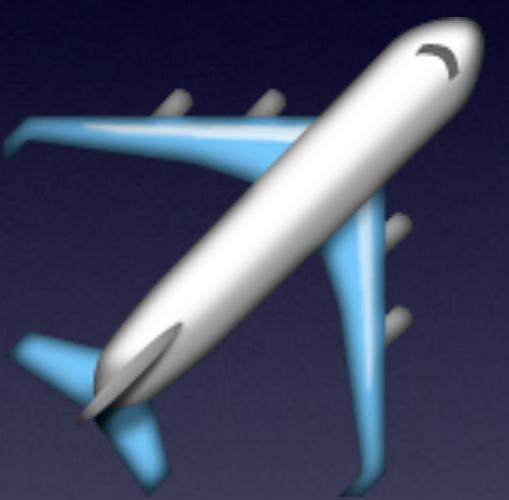


Shift-JIS

This sucks

Unicode!

Unicode!



(planes!)

Basic Multilingual Plane

Code Points

U+0041 (LATIN SMALL
LETTER A)

A

Nº 65
UTF-8 41
General Category Uppercase Letter
Script Latin
Bidirectional Category
Left To Right
Decomposition Type None
East Asian Width Narrow
Font used above DejaVu Serif

U+0041 LATIN CAPITAL LETTER A

U+0041 was added to Unicode in version 1.1. It belongs to the block  Basic Latin in the Basic Multilingual Plane.

This character is a **Uppercase Letter** and is mainly used in the **Latin** script. It is related to its lowercase variant [0061 a](#).

The glyph is **not a composition**. It has a **Narrow** East Asian Width. In bidirectional context it acts as **Left To Right** and is **not mirrored**. The glyph can, under circumstances, be confused with **34 other glyphs**. In text U+0041 behaves as **Alphabetic** regarding line breaks. It has type **Upper** for sentence and **ALetter** for word breaks. The Grapheme Cluster Break is **Any**.

The [Wikipedia](#) has the following information about this codepoint:

A (named a /eɪ/, plural aes) is the first letter and vowel in the ISO basic Latin alphabet. It is similar to the Ancient Greek letter alpha, from which it derives. The upper-case version consists of two more or less vertical lines, joined at the top, and crossed in their middle by a horizontal bar.

Representations

SYSTEM	REPRESENTATION
Nº	65
UTF-8	41
UTF-16	00 41

[show more](#)

Related Characters

0061	00C0	00C1	00C2	00C3	00C4	00C5	0100	0102	0104	01CD	0200	0202	0226	1D2C	1E00	1E9A	1EA0	1EA2	213B	24B6	3373	3380	
a	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	a'	A	A	FAX	Ⓐ	AU	pA	
3381	3382	3383	3384	33DF	FF21	1D400	1D434	1D468	1D49C	1D4D0	1D504	1D538	1D56C	1D5A0	1D5D4	1D608	1D63C	1D670	1F110	1F130			
nA	μA	mA	kA	‰	A	A	A	A	A	A	A	A	A	A	A	A	A	(A)	[A]				

Grapheme

a a a a a A a

Composite
characters

U+0065 U+0301

or

U+00E9

e+ ' => é

é

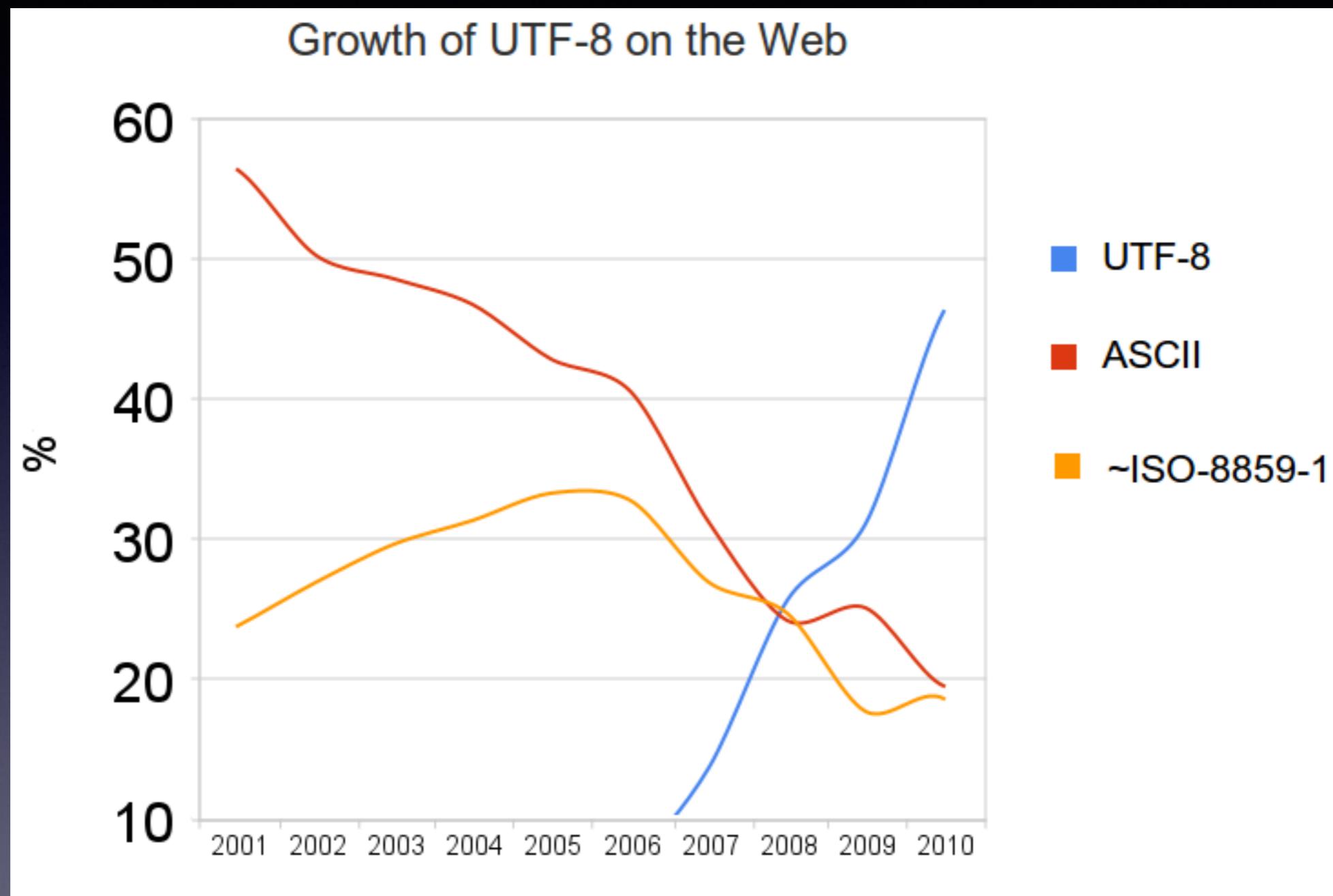
! =

Unicode... is not an
encoding

UTF-32

UCS-2/UTF-16

UTF-8



Source: <http://en.wikipedia.org/wiki/File:UnicodeGrow2b.png>

What does it look
like?

Encoding comparison

Codepoint	Char	ASCII	Latin-1	ISO-8859-15	UTF-8	UTF-16
U+0041	A	0x41	0x41	0x41	0x41	0x00 0x41
U+00C4	Ä	-	0xc4	0xc4	0xc3 0x84	0x00 0xc4
U+20AC	€	-	-	0xa4	0xe3 0x82 0xac	0x20 0xac
U+C218	수	-	-	-	0xec 0x88 0x98	0xc2 0x18

Source: <http://perlgeek.de/en/article/encodings-and-unicode>

Remember: Just
because someone
claims it's UTF-8,
doesn't mean it is