

UCS, Unicode, and ASCII Character Codes

INTRODUCTION

There are 4 general character set codes and many variants. The first is 7 bit (128 characters) ASCII (American Standard Code for Information Interchange) which is the lower 7 bits (high order bit of 0) of the 8 bit (256 character) ASCII code ISO-14962-1997. 8 bit ASCII is the lower 8 bits of the 16 bit (65,536 character) Unicode (high order 8 bits

are 0) [http://www.Unicode.org] Unicode 3.0 was published in February 2000 and is being adopted as ISO/IEC 10646-1:2000 and uses the Unicode Transformation Formats (UTF). Unicode is the lower 16 bits of the 32 bit (4,294,967,296 character) Universal Character Set (UCS) defined by ISO/IEC 10646-1 (the high order 16 bits are 0)

The following ASCII character table is from the Consultative Committee for Space Data Systems (CCSDS):[http://www.ccsds.org/blue_books.html] [http://www.ccsds.org/documents/text/CCSDS-643.0-B-1.txt] document CCSDS 643.0-B-1: *ASCII Encoded English (CCSD0002)*. Blue Book. Issue 1. November 1992. (Reconfirmed June 1998.)

Char: ASCII Character			Dec: Decimal Representation			Hex: Hexadecimal Representation					
Char	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex
NUL	0	00	space	32	20	@	64	40	`	96	60
SOH	1	01	!	33	21	A	65	41	a	97	61
STX	2	02	"	34	22	B	66	42	b	98	62
ETX	3	03	#	35	23	C	67	43	c	99	63
EOT	4	04	\$	36	24	D	68	44	d	100	64
ENQ	5	05	%	37	25	E	69	45	e	101	65
ACK	6	06	&	38	26	F	70	46	f	102	66
BEL	7	07	'	39	27	G	71	47	g	103	67
BS	8	08	(40	28	H	72	48	h	104	68
HT	9	09)	41	29	I	73	49	i	105	69
LF	10	0A	*	42	2A	J	74	4A	j	106	6A
VT	11	0B	+	43	2B	K	75	4B	k	107	6B
FF	12	0C	,	44	2C	L	76	4C	l	108	6C
CR	13	0D	-	45	2D	M	77	4D	m	109	6D
SO	14	0E	.	46	2E	N	78	4E	n	110	6E
SI	15	0F	/	47	2F	O	79	4F	o	111	6F
DLE	16	10	0	48	30	P	80	50	p	112	70
DC1	17	11	1	49	31	Q	81	51	q	113	71
DC2	18	12	2	50	32	R	82	52	r	114	72
DC3	19	13	3	51	33	S	83	53	s	115	73
DC4	20	14	4	52	34	T	84	54	t	116	74
NAK	21	15	5	53	35	U	85	55	u	117	75
SYN	22	16	6	54	36	V	86	56	v	118	76
ETB	23	17	7	55	37	W	87	57	w	119	77
CAN	24	18	8	56	38	X	88	58	x	120	78
EM	25	19	9	57	39	Y	89	59	y	121	79
SUB	26	1A	:	58	3A	Z	90	5A	z	122	7A
ESC	27	1B	;	59	3B	[91	5B	{	123	7B
FS	28	1C	<	60	3C	\	92	5C		124	7C
GS	29	1D	=	61	3D]	93	5D	}	125	7D
RS	30	1E	>	62	3E	^	94	5E	~	126	7E
US	31	1F	?	63	3F	_	95	5F	DEL	127	7F

UCS, Unicode, and ASCII Character Codes

The following table of control character definitions is from [http://www.jimprice.com/jim-asc.htm]

NUL	(null)
SOH	(start of heading)
STX	(start of text)
ETX	(end of text)
EOT	(end of transmission) - Not the same as ETB
ENQ	(enquiry)
ACK	(acknowledge)
BEL	(bell) - Caused teletype machines to ring a bell. Causes a beep in many common terminals and terminal emulation programs.
BS	(backspace) - Moves the cursor (or print head) right to the next tab stop. The spacing of tab stops is dependent on the output device, but is often either 8 or 10.
TAB	(horizontal tab) - Moves the cursor (or print head) right to the next tab stop. The spacing of tab stops is dependent on the output device, but is often either 8 or 10.
LF	(NL line feed, new line) - Moves the cursor (or print head) to a new line. On Unix systems, moves to a new line AND all the way to the left.
VT	(vertical tab)
FF	(form feed) - Advances paper to the top of the next page (if the output device is a printer).
CR	(carriage return) - Moves the cursor all the way to the left, but does not advance to the next line.
SO	(shift out) - Switches output device to alternate character set.
SI	(shift in) - Switches output device back to default character set.
DLE	(data link escape)
DC1	(device control 1)
DC2	(device control 2)
DC3	(device control 3)
DC4	(device control 4)
NAK	(negative acknowledge)
SYN	(synchronous idle)
ETB	(end of transmission block) - Not the same as EOT

CAN	(cancel)
EM	(end of medium)
SUB	(substitute)
ESC	(escape)
FS	(file separator)
GS	(group separator)
RS	(record separator)
US	(unit separator)
DEL	(delete character)

Note to Readers

Updates and More Detailed Descriptions

When using the information in this article, please check the website <http://www.ArchiveBuilders.com> for updates. The version number of this article is just before the page number below. The website also has articles that provide more details on some of the terms and concepts in this article.

Comments

Please let us know how you like this paper, or if you had any questions. What would you like to see in the future? Also, please let us know where you saw this paper. For more, and the most recent version of this article, please visit our web site at <http://www.ArchiveBuilders.com>

Please send your comments via email to SteveGilheany@ArchiveBuilders.com. Tel: +1 310-937-7000. Fax: +1 310-937-7001. Also, please let us know where you saw this article.

Acknowledgements

Reprinted from [Archive Planning](#), Volume 5, number 4, 2001, Archive Builders' analysis newsletter for document management.

See <http://www.ArchiveBuilders.com>

All trademarks are the property of their respective holders.

Note to Editors

Paper 22032v004

We will continue to update these articles as we get comments. Please contact us for the most current version before you publish. Also, please request permission to publish the article. Permission will be given freely for most purposes.

Steve Gilheany
Archive Builders
1209 Manhattan Ave.
Manhattan Beach, CA 90266
Tel: +1 (310) 937-7000 Fax: +1 (310) 937-7001
SteveGilheany@ArchiveBuilders.com

Dividing this Article into Parts for Serialization

If you decide to divide this article into parts please print at least the updates, comments, and acknowledgements sections in each of the parts along with: ' by SteveGilheany@ArchiveBuilders.com'.

Bio

Steve Gilheany, BA in Computer Science, MBA, MLS Specialization in Information Science, CDIA (Certified Document Imaging System Architect), AIIM Master (MIT), and AIIM Laureate (LIT), of Information Technologies, CRM (Certified Records Manager, ARMA) has twenty years experience in document imaging and is a Sr. Systems Engineer at Archive Builders.

Author

Steve Gilheany is a Sr. Systems Engineer at Archive Builders. He has worked in digital document management and document imaging for twenty years.

His experience in the application of document management and document imaging in industry includes: aerospace, banking, manufacturing, natural resources, petroleum refining, transportation, energy, federal, state, and local government, civil engineering, utilities, entertainment, commercial records centers, archives, non-profit development, education, and administrative, engineering, production, legal, and medical records management. At the same time, he has worked in product management for hypertext, for windows based user interface systems, for computer displays, for engineering drawing, letter size, microform, and color scanning, and for xerographic, photographic, newspaper, engineering drawing, and color printing.

In addition, he has nine years of experience in data center operations and database and computer communications systems design, programming, testing, and software configuration management. He has an MLS Specialization in Information Science and an MBA with a concentration in Computer and Information Systems from UCLA, a California Adult Education teaching credential, and a BA in Computer Science from the University of Wisconsin at Madison. His industry certifications include: the CDIA (Certified Document Imaging System Architect) and the AIIM Master (MIT), and AIIM Laureate (LIT), of Information Technologies (from AIIM International, the Association of Information and Image Management, www.AIIM.org), and the CRM (Certified Records Manager) (from the ICRM, the Institute of Certified Records Managers, an affiliate of ARMA International, the Association of Records Managers and Administrators, www.ARMA.org).

Contact:

SteveGilheany@ArchiveBuilders.com
Tel: +1 (310) 937-7000 Fax: +1 (310) 937-7001

For more information, courses, and papers:

<http://www.ArchiveBuilders.com>