Standard ECMA-113
3rd Edition - December 1999

ECMA

Standardizing Information and Communication Systems

8-Bit Single-Byte Coded Graphic Character Sets: Latin/Cyrillic Alphabet

Phone: +41 22 849.60.00 - **Fax:** +41 22 849.60.01 - **URL:** http://www.ecma.ch - **Internet:** helpdesk@ecma.ch

Standard ECMA-113
December 1999

ECMA

Standardizing Information and Communication Systems

8-Bit Single-Byte Coded Graphic Character Sets: Latin/Cyrillic Alphabet

Phone: +41 22 849.60.00 - **Fax:** +41 22 849.60.01 - **URL:** http://www.ecma.ch - **Internet:** helpdesk@ecma.ch

Brief History

The adoption of ECMA-6 (ISO/IEC 646) as the agreed international 7-bit code for information interchange had led to the development of many national, international and application-oriented versions of this code.

These versions had a number of limitations generally inherent to the size of the code:

– they did not provide all graphic characters which were needed;

– for some characters, specially for accented letters, it was necessary to resort to BACKSPACE sequences, which created problems when processing data containing such composite characters;

– interchange among different versions was practically limited to the 82 common graphic characters.

With the advent of 8-bit coding it was possible to increase the number of graphic characters. ISO/IEC 6937, for example, provided a character set covering the requirements of most languages based on the Latin alphabet. This character set, although well suited for text communication, was difficult to use for processing as some graphic characters were represented by one and others by two bit combinations.

Thus the need was recognized for coded graphic character sets, each of which:

– is the same for all users of a given area,

– provides single-byte coding of all graphic characters, thus permitting easy processing,

– takes into account character sets used in the industry.

In 1982 the urgency of the need for an 8-bit single-byte coded character set was recognized in ECMA as well as in ANSI/X3L2 and numerous working papers were exchanged between the two groups. In February 1984 ECMA TC1 submitted to ISO/TC97/SC2 a proposal for such a coded character set. At its meeting of April 1984 SC2 decided to submit to TC97 a proposal for a new item of work for this topic. Technical discussions during and after this meeting led TC1 to adopt the coding scheme proposed by X3L2. International Standard ISO/IEC 8859-1 is based on this joint ANSI/ECMA proposal. ECMA published its corresponding Standard ECMA-94 in March 1985.

After this first publication, the work of ECMA TC1 on further coded graphic character sets has led to the following results:

i. A first Edition, dated June 1986, of a Standard for a Latin/Cyrillic coded graphic character set.

ii. The second Edition of Standard ECMA-94, dated June 1986, comprising four coded graphic character sets for the Latin script, identified as Latin Alphabets No. 1 to No. 4. These alphabets have a number of characters in common, in particular those allocated to columns 02 to 07. They have all been submitted to ISO/IEC JTC 1 - the successor of ISO/TC97 - and are the subject of ISO/IEC 8859, Parts 1 to 4.

iii. A series of ECMA Standards for coded graphic character sets comprising those characters of the Latin Alphabets allocated to columns 02 to 07 and characters of another script for multiple-language applications. These Standards ECMA-114, ECMA-118 and ECMA-121 cover the Arabic, Greek and Hebrew scripts, respectively. They have been submitted to JTC 1 for further processing as ISO/IEC standards and have been published as Part 6, Part 7 and Part 8, respectively, of ISO/IEC 8859.

The 2nd Edition of Standard ECMA-113 superseded the first edition. Indeed, the latter was based on the 1974 version of GOST Standard 19768. In 1987 this standard was revised. As a consequence the 2nd Edition was prepared in co-operation with Russian experts and was brought in complete agreement with the corresponding GOST standard. The corresponding International Standard, ISO/IEC 8859-5:1988 is technically identical with the 2nd Edition of ECMA-113.

In 1999 the 2nd Edition of ISO/IEC 8859-5 has been published, as a technical revision of the 1st Edition of this International Standard. The 3rd Edition of ECMA-113 has been made technically identical with the 2nd Edition of ISO/IEC 8859-5.

This 3rd Edition of Standard ECMA-113 has been adopted by the ECMA General Assembly of December 1999.

Table of contents

1 Scope

This ECMA Standard specifies a set of 191 coded graphic characters identified as the Latin/Cyrillic alphabet.

This set of coded graphic characters is intended for use in data and text processing applications and also for information interchange. The set contains graphic characters used for general purpose applications in typical office environments in at least the following languages:

Bulgarian, Byelorussian, English, Latin, (Slavic) Macedonian, Russian, Serbian and Ukrainian.

NOTE

Two letters recently added to the Ukrainian official alphabet are not included in the character set of this Standard. For a background the CEN/CENELEC/PT004 Report may be consulted (see annex C).

This set of coded graphic characters may be regarded as a version of an 8-bit code according to Standard ECMA-35 or Standard ECMA-43 at level 1.

This Standard may not be used with any other ECMA Standards for 8-bit single-byte coded graphic character sets. If coded characters from more that one ECMA Standard are to be used together, by means of code extension techniques, the equivalent coded character sets from ISO/IEC 10367 should be used instead within a version of Standard ECMA-43 at level 2 or level 3.

The coded characters in this ECMA Standard may be used in conjunction with coded control functions selected from ECMA-48. However, control functions are not used to create composite graphic symbols from two or more graphic characters (see clause 6).

NOTE

This ECMA Standard is not intended for use with Telematic services defined by ITU-T. If information coded according to this ECMA Standard is to be transferred to such services, it will have to conform to the requirements of those services at the access-point.

2 Conformance

2.1 Conformance of information interchange

A coded-character-data-element (CC-data-element) within coded information for interchange is in conformance with this ECMA Standard if all the coded representations of graphic characters within that CC-data-element conform to the requirements of clause 6.

2.2 Conformance of devices

A device is in conformance with this ECMA Standard if it conforms to the requirements of 2.2.1, and either or both of 2.2.2 and 2.2.3. A claim of conformance shall identify the document which contains the description specified in 2.2.1.

2.2.1 Device description

A device that conforms to this ECMA Standard shall be subject of a description that identifies the means by which the user may supply characters to the device, or may recognize them when they are made available to him, as specified respectively in 2.2.2 and 2.2.3.

2.2.2 Originating devices

An originating device shall allow its user to supply any sequence of characters from those specified in clause 6, and shall be capable of transmitting their coded representations within a CC-data-element.

2.2.3 Receiving devices

A receiving device shall be capable of receiving and interpreting any coded representations of characters that are within a CC-data-element, and that conform to clause 6, and shall make the corresponding characters available to its user in such a way that the user can identify them from among those specified there, and can distinguish them from each other.

3 References

ECMA-35	Code Extension Techniques
ECMA-43	8-Bit Coded Character Set Structure and Rules
ECMA-48	Control Functions for Coded Character Sets
ECMA-94	8-Bit Single-Byte Coded Graphic Character Sets - Latin Alphabets No. 1 to No. 4
ECMA-114	8-Bit Single-Byte Coded Graphic Character Sets - Latin/Arabic Alphabet
ECMA-118	8-Bit Single-Byte Coded Graphic Character Sets - Latin/Greek Alphabet
ECMA-121	8-Bit Single-Byte Coded Graphic Character Sets - Latin/Hebrew Alphabet
ECMA-128	8-Bit Single-Byte Coded Graphic Character Sets - Latin alphabet No. 5
ECMA-144	8-Bit Singly-Byte Coded Graphic Character Sets - Latin Alphabet No. 6

4 Definitions

For the purpose of this Standard the following definitions apply.

4.1 bit combination

An ordered set of bits used for the representation of characters.

4.2 byte

A bit string that is operated upon as a unit.

4.3 character

A member of a set of elements used for the organization, control, or representation of data.

4.4 code table

A table showing the characters allocated to each bit combination in a code.

4.5 coded character set; code

A set of unambiguous rules that establishes a character set and the one-to-one relationship between the characters of the set and their bit combinations.

4.6 coded-character-data-element (CC-data-element)

An element of interchanged information that is specified to consist of a sequence of coded representations of characters, in accordance with one or more identified standards for coded character sets.

4.7 graphic character

A character, other than a control function, that has a visual representation normally hand-written, printed or displayed, and that has a coded representation consisting of one or more bit combinations.

NOTE

In this Standard a single bit combination is used to represent each character.

4.8 graphic symbol

A visual representation of a graphic character or of a control function.

4.9 position

That part of a code table identified by its column and row co-ordinates.

5 Notation, code table and names

5.1 Notation

The bits of the bit combinations of the 8-bit code are identified by b_8, b_7, b_6, b_5, b_4, b_3, b_2 and b_1, where b_8 is the highest-order, or most-significant bit and b_1 is the lowest-order, or least-significant bit.

The bit combinations may be interpreted to represent numbers in binary notation by attributing the following weights to the individual bits:

Bit	b_8	b_7	b_6	b_5	b_4	b_3	b_2	b_1
Weight	128	64	32	16	8	4	2	1

Using these weights, the bit combinations are identified by notations of the form xx/yy, where xx and yy are numbers in the range 00 to 15. The correspondence between the notations of the form xx/yy and the bit combinations consisting of the bits b_8 to b_1 is as follows:

– xx is the number represented by b_8, b_7, b_6 and b_5 where these bits are given the weights 8, 4, 2, and 1, respectively.

– yy is the number represented by b_4, b_3, b_2 and b_1 where these bits are given the weights 8, 4, 2, and 1, respectively.

The bit combinations are also identified by notations of the form hk, where h and k are numbers in the range 0 to F in hexadecimal notation. The number h is the same as the number xx described above, and the number k the same as the number yy described above.

5.2 Layout of the code table

An 8-bit code table consists of 256 positions arranged in 16 columns and 16 rows. The columns and the rows are numbered 00 to 15. In hexadecimal notation the columns and the rows are numbered 0 to F.

The code table positions are identified by notations of the form xx/yy, where xx is the column number and yy is the row number. The column and row numbers are shown at the top and left edges of the table, respectively. The code table positions are also identified by notations of the form hk, where h is the column number and k is the row number in hexadecimal notation. The column and row numbers are shown at the bottom and right edges of the table, respectively.

The positions of the code table are in one-to-one correspondence with the bit combinations of the code. The notation of a code table position, of the form xx/yy, or of the form hk, is the same as that of the corresponding bit combination.

5.3 Names and meanings.

This ECMA Standard assigns a unique name and a unique identifier to each graphic character. These names and identifiers have been taken from ISO/IEC 10646-1. This ECMA Standard also specifies an acronym for each of the characters SPACE, NO-BREAK SPACE and SOFT HYPHEN. For acronyms only Latin capital letters A to Z are used. It is intended that the acronyms be retained in all translations of the text.

Except for SPACE (SP), NO-BREAK SPACE (NBSP) and SOFT HYPHEN (SHY), this ECMA Standard does not define and does not restrict the meanings of graphic characters.

This ECMA Standard specifies a graphic symbol for each graphic character. This symbol is shown in the corresponding position of the code table. However, this Standard does not specify a particular style or font design for imaging graphic characters.

5.3.1 SPACE (SP)

A graphic character the visual representation of which consists of the absence of a graphic symbol.

5.3.2 NO-BREAK SPACE (NBSP)

A graphic character the visual representation of which consists of the absence of a graphic symbol, for use when a line break is to be prevented in the text as presented.

5.3.3 SOFT HYPHEN (SHY)

A graphic character that is imaged by a graphic symbol identical with, or similar to, that representing HYPHEN, for use when a line break has been established within a word.

6 Specification of the coded character set

This ECMA Standard specifies 191 characters allocated to the bit combinations of the code table (table 2). None of these characters are combining characters.

NOTE

Combining characters are described in ECMA-35, subclause 6.3.3.

Control functions, such as BACKSPACE or CARRIAGE RETURN, shall not be used to create composite graphic symbols, which are made up from the graphic representations of two or more characters.

6.1 Characters of the set and their coded representation

See table 1.

Table 1 - Character set, coded representation

Bit combina-tion	Hex	Identifier	Name
02/00	20	U+0020	SPACE
02/01	21	U+0021	EXCLAMATION MARK
02/02	22	U+0022	QUOTATION MARK
02/03	23	U+0023	NUMBER SIGN
02/04	24	U+0024	DOLLAR SIGN
02/05	25	U+0025	PERCENT SIGN
02/06	26	U+0026	AMPERSAND
02/07	27	U+0027	APOSTROPHE
02/08	28	U+0028	LEFT PARENTHESIS
02/09	29	U+0029	RIGHT PARENTHESIS
02/10	2A	U+002A	ASTERISK
02/11	2B	U+002B	PLUS SIGN
02/12	2C	U+002C	COMMA
02/13	2D	U+002D	HYPHEN-MINUS
02/14	2E	U+002E	FULL STOP
02/15	2F	U+002F	SOLIDUS
03/00	30	U+0030	DIGIT ZERO
03/01	31	U+0031	DIGIT ONE
03/02	32	U+0032	DIGIT TWO
03/03	33	U+0033	DIGIT THREE
03/04	34	U+0034	DIGIT FOUR
03/05	35	U+0035	DIGIT FIVE
03/06	36	U+0036	DIGIT SIX
03/07	37	U+0037	DIGIT SEVEN
03/08	38	U+0038	DIGIT EIGHT
03/09	39	U+0039	DIGIT NINE
03/10	3A	U+003A	COLON
03/11	3B	U+003B	SEMICOLON
03/12	3C	U+003C	LESS-THAN SIGN
03/13	3D	U+003D	EQUALS SIGN
03/14	3E	U+003E	GREATER-THAN SIGN
03/15	3F	U+003F	QUESTION MARK
04/00	40	U+0040	COMMERCIAL AT
04/01	41	U+0041	LATIN CAPITAL LETTER A
04/02	42	U+0042	LATIN CAPITAL LETTER B
04/03	43	U+0043	LATIN CAPITAL LETTER C
04/04	44	U+0044	LATIN CAPITAL LETTER D
04/05	45	U+0045	LATIN CAPITAL LETTER E
04/06	46	U+0046	LATIN CAPITAL LETTER F

Bit combina-tion	Hex	Identifier	Name
04/07	47	U+0047	LATIN CAPITAL LETTER G
04/08	48	U+0048	LATIN CAPITAL LETTER H
04/09	49	U+0049	LATIN CAPITAL LETTER I
04/10	4A	U+004A	LATIN CAPITAL LETTER J
04/11	4B	U+004B	LATIN CAPITAL LETTER K
04/12	4C	U+004C	LATIN CAPITAL LETTER L
04/13	4D	U+004D	LATIN CAPITAL LETTER M
04/14	4E	U+004E	LATIN CAPITAL LETTER N
04/15	4F	U+004F	LATIN CAPITAL LETTER O
05/00	50	U+0050	LATIN CAPITAL LETTER P
05/01	51	U+0051	LATIN CAPITAL LETTER Q
05/02	52	U+0052	LATIN CAPITAL LETTER R
05/03	53	U+0053	LATIN CAPITAL LETTER S
05/04	54	U+0054	LATIN CAPITAL LETTER T
05/05	55	U+0055	LATIN CAPITAL LETTER U
05/06	56	U+0056	LATIN CAPITAL LETTER V
05/07	57	U+0057	LATIN CAPITAL LETTER W
05/08	58	U+0058	LATIN CAPITAL LETTER X
05/09	59	U+0059	LATIN CAPITAL LETTER Y
05/10	5A	U+005A	LATIN CAPITAL LETTER Z
05/11	5B	U+005B	LEFT SQUARE BRACKET
05/12	5C	U+005C	REVERSE SOLIDUS
05/13	5D	U+005D	RIGHT SQUARE BRACKET
05/14	5E	U+005E	CIRCUMFLEX ACCENT
05/15	5F	U+005F	LOW LINE
06/00	60	U+0060	GRAVE ACCENT
06/01	61	U+0061	LATIN SMALL LETTER A
06/02	62	U+0062	LATIN SMALL LETTER B
06/03	63	U+0063	LATIN SMALL LETTER C
06/04	64	U+0064	LATIN SMALL LETTER D
06/05	65	U+0065	LATIN SMALL LETTER E
06/06	66	U+0066	LATIN SMALL LETTER F
06/07	67	U+0067	LATIN SMALL LETTER G
06/08	68	U+0068	LATIN SMALL LETTER H
06/09	69	U+0069	LATIN SMALL LETTER I
06/10	6A	U+006A	LATIN SMALL LETTER J
06/11	6B	U+006B	LATIN SMALL LETTER K
06/12	6C	U+006C	LATIN SMALL LETTER L
06/13	6D	U+006D	LATIN SMALL LETTER M
06/14	6E	U+006E	LATIN SMALL LETTER N
06/15	6F	U+006F	LATIN SMALL LETTER O
07/00	70	U+0070	LATIN SMALL LETTER P
07/01	71	U+0071	LATIN SMALL LETTER Q
07/02	72	U+0072	LATIN SMALL LETTER R
07/03	73	U+0073	LATIN SMALL LETTER S
07/04	74	U+0074	LATIN SMALL LETTER T
07/05	75	U+0075	LATIN SMALL LETTER U
07/06	76	U+0076	LATIN SMALL LETTER V
07/07	77	U+0077	LATIN SMALL LETTER W
07/08	78	U+0078	LATIN SMALL LETTER X
07/09	79	U+0079	LATIN SMALL LETTER Y

Bit combina-tion	Hex	Identifier	Name
07/10	7A	U+007A	LATIN SMALL LETTER Z
07/11	7B	U+007B	LEFT CURLY BRACKET
07/12	7C	U+007C	VERTICAL LINE
07/13	7D	U+007D	RIGHT CURLY BRACKET
07/14	7E	U+007E	TILDE
10/00	A0	U+00A0	NO-BREAK SPACE
10/01	A1	U+0401	CYRILLIC CAPITAL LETTER IO
10/02	A2	U+0402	CYRILLIC CAPITAL LETTER DJE
10/03	A3	U+0403	CYRILLIC CAPITAL LETTER GJE
10/04	A4	U+0404	CYRILLIC CAPITAL LETTER UKRANIAN IE
10/05	A5	U+0405	CYRILLIC CAPITAL LETTER DZE
10/06	A6	U+0406	CYRILLIC CAPITAL LETTER BYELORUSSIAN-UKRANIAN I
10/07	A7	U+0407	CYRILLIC CAPITAL LETTER YI
10/08	A8	U+0408	CYRILLIC CAPITAL LETTER JE
10/09	A9	U+0409	CYRILLIC CAPITAL LETTER LJE
10/10	AA	U+040A	CYRILLIC CAPITAL LETTER NJE
10/11	AB	U+040B	CYRILLIC CAPITAL LETTER TSHE
10/12	AC	U+040C	CYRILLIC CAPITAL LETTER KJE
10/13	AD	U+040D	SOFT HYPHEN
10/14	AE	U+040E	CYRILLIC CAPITAL LETTER SHORT U
10/15	AF	U+040F	CYRILLIC CAPITAL LETTER DZHE
11/00	B0	U+0410	CYRILLIC CAPITAL LETTER A
11/01	B1	U+0411	CYRILLIC CAPITAL LETTER BE
11/02	B2	U+0412	CYRILLIC CAPITAL LETTER VE
11/03	B3	U+0413	CYRILLIC CAPITAL LETTER GHE
11/04	B4	U+0414	CYRILLIC CAPITAL LETTER DE
11/05	B5	U+0415	CYRILLIC CAPITAL LETTER IE
11/06	B6	U+0416	CYRILLIC CAPITAL LETTER ZHE
11/07	B7	U+0417	CYRILLIC CAPITAL LETTER ZE
11/08	B8	U+0418	CYRILLIC CAPITAL LETTER I
11/09	B9	U+0419	CYRILLIC CAPITAL LETTER SHORT I
11/10	BA	U+041A	CYRILLIC CAPITAL LETTER KA
11/11	BB	U+041B	CYRILLIC CAPITAL LETTER EL
11/12	BC	U+041C	CYRILLIC CAPITAL LETTER EM
11/13	BD	U+041D	CYRILLIC CAPITAL LETTER EN
11/14	BE	U+041E	CYRILLIC CAPITAL LETTER O
11/15	BF	U+041F	CYRILLIC CAPITAL LETTER PE
12/00	C0	U+0420	CYRILLIC CAPITAL LETTER ER
12/01	C1	U+0421	CYRILLIC CAPITAL LETTER ES
12/02	C2	U+0422	CYRILLIC CAPITAL LETTER TE
12/03	C3	U+0423	CYRILLIC CAPITAL LETTER U
12/04	C4	U+0424	CYRILLIC CAPITAL LETTER EF
12/05	C5	U+0425	CYRILLIC CAPITAL LETTER HA
12/06	C6	U+0426	CYRILLIC CAPITAL LETTER TSE
12/07	C7	U+0427	CYRILLIC CAPITAL LETTER CHE
12/08	C8	U+0428	CYRILLIC CAPITAL LETTER SHA
12/09	C9	U+0429	CYRILLIC CAPITAL LETTER SHCHA
12/10	CA	U+042A	CYRILLIC CAPITAL LETTER HARD SIGN
12/11	CB	U+042B	CYRILLIC CAPITAL LETTER YERU
12/12	CC	U+042C	CYRILLIC CAPITAL LETTER SOFT SIGN
12/13	CD	U+042D	CYRILLIC CAPITAL LETTER E

Bit combina-tion	Hex	Identifier	Name
12/14	CE	U+042E	CYRILLIC CAPITAL LETTER YU
12/15	CF	U+042F	CYRILLIC CAPITAL LETTER YA
13/00	D0	U+0430	CYRILLIC SMALL LETTER A
13/01	D1	U+0431	CYRILLIC SMALL LETTER BE
13/02	D2	U+0432	CYRILLIC SMALL LETTER VE
13/03	D3	U+0433	CYRILLIC SMALL LETTER GHE
13/04	D4	U+0434	CYRILLIC SMALL LETTER DE
13/05	D5	U+0435	CYRILLIC SMALL LETTER IE
13/06	D6	U+0436	CYRILLIC SMALL LETTER ZHE
13/07	D7	U+0437	CYRILLIC SMALL LETTER ZE
13/08	D8	U+0438	CYRILLIC SMALL LETTER I
13/09	D9	U+0439	CYRILLIC SMALL LETTER SHORT I
13/10	DA	U+043A	CYRILLIC SMALL LETTER KA
13/11	DB	U+043B	CYRILLIC SMALL LETTER EL
13/12	DC	U+043C	CYRILLIC SMALL LETTER EM
13/13	DD	U+043D	CYRILLIC SMALL LETTER EN
13/14	DE	U+043E	CYRILLIC SMALL LETTER O
13/15	DF	U+043F	CYRILLIC SMALL LETTER PE
14/00	E0	U+0440	CYRILLIC SMALL LETTER ER
14/01	E1	U+0441	CYRILLIC SMALL LETTER ES
14/02	E2	U+0442	CYRILLIC SMALL LETTER TE
14/03	E3	U+0443	CYRILLIC SMALL LETTER U
14/04	E4	U+0444	CYRILLIC SMALL LETTER EF
14/05	E5	U+0445	CYRILLIC SMALL LETTER HA
14/06	E6	U+0446	CYRILLIC SMALL LETTER TSE
14/07	E7	U+0447	CYRILLIC SMALL LETTER CHE
14/08	E8	U+0448	CYRILLIC SMALL LETTER SHA
14/09	E9	U+0449	CYRILLIC SMALL LETTER SHCHA
14/10	EA	U+044A	CYRILLIC SMALL LETTER HARD SIGN
14/11	EB	U+044B	CYRILLIC SMALL LETTER YERU
14/12	EC	U+044C	CYRILLIC SMALL LETTER SOFT SIGN
14/13	ED	U+044D	CYRILLIC SMALL LETTER E
14/14	EE	U+044E	CYRILLIC SMALL LETTER YU
14/15	EF	U+044F	CYRILLIC SMALL LETTER YA
15/00	F0	U+2116	NUMERO SIGN
15/01	F1	U+0451	CYRILLIC SMALL LETTER IO
15/02	F2	U+0452	CYRILLIC SMALL LETTER DJE
15/03	F3	U+0453	CYRILLIC SMALL LETTER GJE
15/04	F4	U+0454	CYRILLIC SMALL LETTER UKRANIAN IE
15/05	F5	U+0455	CYRILLIC SMALL LETTER DZE
15/06	F6	U+0456	CYRILLIC SMALL LETTER BYELORUSSIAN-UKRANIAN I
15/07	F7	U+0457	CYRILLIC SMALL LETTER YI
15/08	F8	U+0458	CYRILLIC SMALL LETTER JE
15/09	F9	U+0459	CYRILLIC SMALL LETTER LJE
15/10	FA	U+045A	CYRILLIC SMALL LETTER NJE
15/11	FB	U+045B	CYRILLIC SMALL LETTER TSHE
15/12	FC	U+045C	CYRILLIC SMALL LETTER KJE
15/13	FD	U+00A7	SECTION SIGN
15/14	FE	U+045E	CYRILLIC SMALL LETTER SHORT U
15/15	FF	U+045F	CYRILLIC SMALL LETTER DZHE

6.2 Code table

For each character in the set the code table (table 2) shows a graphic symbol at the position in the code table corresponding to the bit combination specified in table 1.

The shaded positions in the code table correspond to bit combinations that do not represent graphic characters. Their use is outside the scope of this Standard; it is specified in other Standards, for example in Standard ECMA-48.

Table 2 - Code table of Latin/Cyrillic alphabet

b_4	b_3	b_2	b_1		b_8 0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
					b_7 0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
					b_6 0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
					b_5 0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
					00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
0	0	0	0	00			SP	0	@	P	`	p			NBSP	А		а	р	№	0
0	0	0	1	01			!	1	A	Q	a	q			Ё	Б	С	б	с	ё	1
0	0	1	0	02			"	2	B	R	b	r			Ђ	В	Т	в	т	ђ	2
0	0	1	1	03			#	3	C	S	c	s			Ѓ	Г	У	г	у	ѓ	3
0	1	0	0	04			$	4	D	T	d	t			Є	Д	Ф	д	ф	є	4
0	1	0	1	05			%	5	E	U	e	u			S	Е	Х	е	х	s	5
0	1	1	0	06			&	6	F	V	f	v			І	Ж	Ц	ж	ц	і	6
0	1	1	1	07			'	7	G	W	g	w			Ї	З	Ч	з	ч	ї	7
1	0	0	0	08			(	8	H	X	h	x			J	И	Ш	и	ш	j	8
1	0	0	1	09			)	9	I	Y	i	y			Љ	Й	Щ	й	щ	љ	9
1	0	1	0	10			*	:	J	Z	j	z			Њ	К	Ъ	к	ъ	њ	A
1	0	1	1	11			+	;	K	[	k	{			Ћ	Л	Ы	л	ы	ћ	B
1	1	0	0	12			,	<	L	\	l	|			Ќ	М	Ь	м	ь	ќ	C
1	1	0	1	13			-	=	M	]	m	}			SHY	Н	Э	н	э	§	D
1	1	1	0	14			.	>	N	^	n	~			Ў	О	Ю	о	ю	ў	E
1	1	1	1	15			/	?	O	_	o				Џ	П	Я	п	я	џ	F
					0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	Hex

99-0086-A

7 Identification of the character set

7.1 Identification according to ECMA-35 and ECMA-43

The graphic characters of this ECMA Standard constitute a single coded character set. However, in accordance with ECMA-35 and ECMA-43 the code table of this ECMA Standard may be considered to consist of the following components:

– The character SPACE represented by bit combination 02/00;

– a 94-character G0 graphic character set represented by bit combinations 02/01 to 07/14;

– a 96-character G1 graphic character set represented by bit combinations 10/00 to 15/15.

When the identification methods of ECMA-35 or ECMA-43 are used, this ECMA Standard shall be identified by the following pair of designation functions:

GZD4 04/02 (ESC 02/08 04/02)

G1D6 04/12 (ESC 02/13 04/12)

NOTE

The corresponding escape sequences are shown in parentheses.

7.2 Identification using the ISO International register of coded character sets to be used with escape sequences

According to 7.1 above the character set of this ECMA Standard may be considered to consist of the character SPACE, a 94-character G0 graphic character set, and a 96-character G1 graphic character set. The G0 and G1 graphic character sets may be identified by the use of the Registration Numbers from the ISO International register of coded character sets to be used with escape sequences.

When these Registration Numbers are used this ECMA Standard shall be identified by the following pair of registration numbers:

– G0 graphic character set ISO-IR 6

– G1 graphic character set ISO-IR 144

Annex A
(informative)

Coverage of languages

A.1 Languages of European origin written in Latin script

The following ECMA Standards specify coded character sets which comprise various different selections of characters based on the Latin alphabet. These sets are identified by the numbers 1 to 6 as shown:

ECMA-94 Latin alphabets No. 1 to 4
ECMA-128 Latin alphabet No. 5
ECMA-144 Latin alphabet No. 6

Table A.1 - Language coverage

Language	Covered by alphabet(s)	Language	Covered by alphabet(s)	Language	Covered by alphabet(s)
Albania	1 2 5	Frisian	1 5	Norwegian	1 4 5 6
Basque	1 5	Galician	1 5	Polish	2
Breton	1 5	German	1 2 3 4 5 6	Portuguese	1 3 5
Catalan	1 5	Greenlandic	1 4 5 6	Rhaeto-Romanic	1 5
Croat	2	Hungarian	2	Romanian	2
Czech	2	Icelandic	1 6	Sámi	4 6
Danish	1 4 5 6	Irish Gaelic	1 5 6	Scottish Gaelic	1 5
Dutch	1 5	*(new orthography)*		Slovak	2
English	1 2 3 4 5 6	Italian	1 3 5	Slovene	2 4 6
Esperanto	3	Latin	1 2 3 4 5 6	Serbian	2
Estonian	4 6	Latvian	4	Spanish	1 5
Faroese	1 6	Lithuanian	4 6	Swedish	1 4 5 6
Finnish	1 4 5 6	Luxemburgish	1 5	Turkish	(3) 5
French	(1) (3) (5)	Maltese	3		

NOTES

1. *The list of languages in table A.1 is not exhaustive. It shows the languages that are included in the Scope clause of each of the ECMA Standardsfor the Latin alphabets.*
2. *For writing French, three characters (Œ, œ, Ÿ) not specified in Latin alphabets No. 1, 3 and 5, are also needed.*
3. *The various Sámi languages use partly differing orthographies. The character sets in Latin alphabets No. 4 and No. 6 cover the requirements of the Sámi languages most commonly used in Finland, Norway and Sweden. For the Skolt Sámi language used in Finland and Norway additional characters are needed.*
4. *There are several official written languages outside Europe that are covered by Latin alphabet No. 1. Examples are Indonesian/Malay, Tagalog (Philippines), Swahili, Afrikaans.*
5. *Use of Latin alphabet No. 3 for Turkish is deprecated.*

A.2 Languages written in non-Latin scripts

The following standards specify coded character sets which include graphic characters from alphabets other than the Latin alphabet:

ECMA-113 Latin/Cyrillic alphabet
ECMA-114 Latin/Arabic alphabet
ECMA-118 Latin/Greek alphabet
ECMA-121 Latin/Hebrew alphabet

The following official and regional languages are covered by these alphabets:

The Cyrillic characters included in this ECMA Standard cover Bulgarian, Byelorussian, (Slavic) Macedonian, Russian, Serbian and Ukrainian (as written up to 1990, see also the Scope of this ECMA Standard).

The Arabic characters included in ECMA-114 cover Arabic. The Greek characters included in ECMA-118 cover Greek (*monotonikó* orthography). The Hebrew characters included in ECMA-121 cover Hebrew.

Annex B
(informative)

Main differences between the second edition and this third edition of ECMA-113

B.1 The names of the graphic characters have been amended where necessary to align them with the names of the characters adopted for all standards on coded character sets developed under the responsibility of ISO/IEC JTC 1. For each character the short identifiers specified in ISO/IEC 10646-1, Amendment 9, have been added to table 1.

B.2 The new style of conformance clause, adopted for all standards on coded character sets, has been introduced.

B.3 Object identifiers conforming to Abstract Syntax Notation One are specified in annex D for the character set, and the corresponding coded representations of this ECMA Standard.

Registration numbers from the International register of coded character sets to be used with escape sequences have been included as an additional method of identifying the coded character set of this ECMA Standard.

B.4 A new annex A has been added that identifies the coverage of languages by the Standards for the Latin alphabets.

B.5 Various editorial adjustments and clarifications have been made to the text of the Standard. The hexadecimal equivalents of the bit combinations have been added to tables 1 and 2.

B.6 Annex C, Bibliography, and annex D, Identification according to ISO/IEC 8824-1, have been added.

Annex C
(informative)

Bibliography

ECMA-48 Control Functions for Coded Character Sets, 5th Edition (June 1991)

ISO/IEC 10367:1991 Information technology - Standardized coded graphic character sets for use in 8-bit codes

ISO/IEC 10646-1:1993 Information technology - Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and Basic Multilingual Plane

ISO International register of coded character sets to be used with escape sequences.

CEN/CENELEC IT/PT004, Report from the project team on Definition of a Cyrillic primary set of graphic characters (CEN, Brussels, July 1992)

Annex D
(informative)

Identification according to ISO/IEC 8824-1 (ASN.1)

In the terminology of ISO/IEC 8824-1 the character set of part of ISO/IEC 8859-5 (ECMA-113) and the corresponding coded representations are distinct, and are known as the "character abstract syntax" and the "character transfer syntax", respectively.

When the identification methods of ISO/IEC 8824-1 are used, ISO/IEC 8859-5 shall be identified by the following object identifiers:

– character set
  {iso standard 8859 5 abstract-syntax (1)}

– coded representations
  {iso standard 8859 5 transfer-syntax (0)}

The corresponding object descriptors shall be:

– character set "ISO 8859 part 5 repertoire"

– coded representations "ISO 8859 part 5 code".

Free printed copies can be ordered from:

ECMA
114 Rue du Rhône
CH-1204 Geneva
Switzerland

Fax: +41 22 849.60.01
Internet: documents@ecma.ch

Files of this Standard can be freely downloaded from the ECMA web site (www.ecma.ch). This site gives full information on ECMA, ECMA activities, ECMA Standards and Technical Reports.

ECMA

114 Rue du Rhône
CH-1204 Geneva
Switzerland

See inside cover page for obtaining further soft or hard copies.