

Формат ASCII для обмена данными через сеть

ASCII format for Network Interchange

Для конкретности предлагается использовать стандартный 7-битовый код ASCII, встраиваемый в 8-битовые байты, старший бит которых всегда имеет значение 0. Это позволяет воспользоваться стандартным кодом из USAS X3, 4-1968. Данный код будет использоваться прежде всего для обмена данными между хостами. Символы прерывания будут определяться принимающим хостом. Например, SRI использует точку «.» (ASCII X'2E' или 2/14) в качестве символа завершения строки, а UCLA использует X'OD' или 0/13 (возврат каретки).

Стандартный код США для обмена информацией.

1. Область применения

Этот кодируемый набор символов будет использоваться для информационного обмена общего назначения между системами обработки информации, коммуникационным оборудованием и связанными с ними устройствами.

2. Стандартный код

					----->							
Б \ b7					0	0	0	0	1	1	1	1
и \ b6					0	0	1	1	0	0	1	1
т \ b5					0	1	0	1	0	1	0	1
ы					----->							
Столбец->					0	1	2	3	4	5	6	7
b4	b3	b2	b1	Строка								
0	0	0	0	0	NUL	DLE	SP	0	@	P		p
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q
0	0	1	0	2	STX	DC2	"	2	B	R	b	r
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s
0	1	0	0	4	EOT	DC4	\$	4	D	T	d	t
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w
1	0	0	0	8	BS	CAN	(8	H	X	h	x
1	0	0	1	9	HT	EM)	9	I	Y	i	y
1	0	1	0	10	LF	SUB	*	:	J	Z	j	z
1	0	1	1	11	VT	ESC	+	;	K	[k	{
1	1	0	0	12	FF	FS	,	<	L	\	l	
1	1	0	1	13	CR	GS	-	=	M]	m	}
1	1	1	0	14	SO	RS	.	>	N	^	n	~
1	1	1	1	15	SI	US	/	?	O	_	o	DEL

3. Представление символов и идентификация кода

Стандартное 7-битовое представление символа K, где b7 указывает старший, а b1 — младший бит, как показано ниже.

Пример. Битовое представление символа K (столбец 4, строка 11) будет иметь вид

```

b7 b6 b5 b4 b3 b2 b1
1 0 0 1 0 1 1
    
```

Положение символа K в кодовой таблице можно также указать в форме «столбец 4, строка 11» или просто 4/11. десятичный эквивалент двоичного числа, формируемого битами b7, b6 и b5, определяет номер столбца, а десятичный эквивалент двоичного числа, образуемого битами b4, b3, b2 и b1, - номер строки в таблице кодов.

Для идентификации этого стандартного кода используется аббревиатура ASCII или USASCII.

Обозначение ASCII (произносится as'-key) или USASCII (произносится you-sas'-key) обычно следует применять для кода, предписанного последней версией стандарта. Для явного обозначения конкретной (возможно, предварительной) версии следует добавлять к аббревиатуре две последние цифры года — например, ASCII 63 или USASCII 63.

4. Описания символов

4.1 Управляющие символы

Обозначение	Значение	Перевод
NUL	Null	Пустой символ
SOH ¹	Start of Heading	Начало заголовка
STX ¹	Start of Text	Начало текста
ETX ¹	End of Text	Конец текста
EOT ¹	End of Transmission	Завершение передачи
ENQ ¹	Enquiry	Запрос
ACK ¹	Acknowledge	Подтверждение
BEL	Bell	Звонок (звуковой или иной сигнал)
BS ²	Backspace	Забой символа
HT ²	Horizontal Tabulation	Горизонтальная табуляция (пропуск перфокарты)
LF ²	Line Feed	Перевод строки
VT ²	Vertical Tabulation	Вертикальная табуляция
FF ²	Form Feed	Новая страница
CR ²	Carriage Return	Возврат каретки
SO	Shift Out	Начало нестандартной трактовки символов
SI	Shift In	Возврат к стандартной трактовке символов
DLE ¹	Data Link Escape	Особая трактовка символов
DC1	Device Control 1	Управление устройством 1
DC2	Device Control 2	Управление устройством 2
DC3	Device Control 3	Управление устройством 3
DC4	Device Control 4	Управление устройством 4 (стоп)
NAK ¹	Negative Acknowledge	Негативное подтверждение
SYN ¹	Synchronous Idle	Синхронизация при бездействии
ETB ¹	End of Transmission Block	Конец блока передачи
CAN	Cancel	Отказ
EM	End of Medium	Конец среды (носителя)
SUB	Substitute	Замена
ESC	Escape	
FS ³	File Separator	Разделитель файлов
GS ³	Group Separator	Разделитель групп
RS ³	Record Separator	Разделитель записей
US ³	Unit Separator	Разделитель блоков
DEL ⁴	Delete	Удаление

¹Управление коммуникациями (CC). См. параграф 5.1.

²Форматирование (FE). См. параграф 5.1.

³Разделитель (IS). См. параграф 5.1.

⁴Строго говоря, символ DEL не относится к управляющим (см. параграф 5.2).

4.2 Символы графики

Строка/столбец	Символ	Название	Значение
2/0	SP	Space	Пробел (обычно не печатается)
2/1	!	Exclamation Point	Знак восклицания
2/2 ¹	"	Quotation Marks	Кавычка
2/3 ²	#	Number Sign	Знак номера ³
2/4	\$	Dollar Sign	Знак доллара
2/5	%	Percent	Процент
2/6	&	Ampersand	Амперсанд
2/7	'	Apostrophe	Апостроф (одинарная закрывающая кавычка, знак ударения ¹)
2/8	(Opening Parenthesis	Открывающая скобка
2/9)	Closing Parenthesis	Закрывающая скобка
2/10	*	Asterisk	Звездочка
2/11	+	Plus	Плюс
2/12	,	Comma	Запятая (седиль ¹)
2/13	-	Hyphen (Minus)	Дефис (минус)
2/14	.	Period (Decimal Point)	Точка (десятичный разделитель)
2/15	/	Slant	Дробная черта
3/10	:	Colon	Двоеточие
3/11	;	Semicolon	Точка с запятой
3/12	<	Less Than	Знак «меньше»
3/13	=	Equals	Знак равенства
3/14	>	Greater Than	Знак «больше»
3/15	?	Question Mark	Знак вопроса
4/0 ²	@	Commercial At	
5/11 ²	[Opening Bracket	Открывающая квадратная скобка
5/12 ²	\	Reverse Slant	Обратная дробная черта
5/13 ²]	Closing Bracket	Закрывающая квадратная скобка
5/14 ^{1,2}	^	Circumflex	Циркумфлекс
5/15	_	Underline	Подчеркивание
6/0 ^{1,2}	`	Grave Accent	Гравис (открывающая одинарная кавычка)
7/11 ²	{	Opening Brace	Открывающая фигурная скобка
7/12 ²		Vertical Line	Вертикальная линия
7/13 ²	}	Closing Brace	Закрывающая фигурная скобка
7/14 ²	~	Overline	Оверлайн (тильда ¹ , общее ударение ¹)

5. Определения

5.1 Общие

(CC) *Communication Control* — управление коммуникациями

Функциональные символы, предназначенные для управления или облегчения передачи информации через коммуникационные сети.

(FE) *Format Effector* - форматирование

Функциональные символы, управляющие размещением символов на устройствах печати и отображения.

(IS) *Information Separator* - разделитель

Символы, используемые для логического разделения информации на те или иные блоки. Группа разделителей включает четыре символа, образующих иерархию порядка применения.

¹Использование символов 2/2, 2/7, 2/12, 5/14, 6/0 и 7/14 в качестве диакритических знаков, описано в Приложении A5.2 к X3.4-1968 (в оригинале документ X3.4-1968 ошибочно не указан, см. https://www.rfc-editor.org/errata_search.php?eid=4217, прим. перев.).

²Не следует использовать при международном обмене без согласования между отправителем и получателем, см. Приложение B4 к X3.4-1968 (в оригинале документ X3.4-1968 ошибочно не указан, см. https://www.rfc-editor.org/errata_search.php?eid=4217, прим. перев.).

³В Приложениях, где не требуется символ номера (#), взамен может применяться символ фунта стерлингов.

5.2 Управляющие символы

NUL (Null) — пустой символ

Символ, все биты которого имеют значение 0. Такие символы могут применяться для заполнения времени или пространства в среде (носителе).

SOH (Start of Heading) — начало заголовка

Символ управления коммуникациями, который обозначает начало последовательности символов, задающий понятный оборудованию адрес или маршрутную информацию. Такие последовательности называют заголовками (heading). Для обозначения конца заголовка служит символ STX.

STX (Start of Text) — начало текста

Символ управления коммуникациями, предшествующий последовательности символов, которая должна трактоваться как единое целое и полностью передаваться конечному получателю. Такую последовательность символов называют текстом (text). Символ STX может использоваться для завершения последовательности символов, начинающейся с SOH.

ETX (End of Text) — завершение текста

Символ управления коммуникациями, используемый для обозначения конца последовательности символов, начинающейся с STX и передаваемой как единое целое.

EOT (End of Transmission) — завершение передачи

Символ управления коммуникациями, используемый для индикации завершения передачи, которая может содержать один или множество тестов и связанных с ними заголовков.

ENQ (Enquiry) — запрос

Символ управления коммуникациями, используемый в коммуникационных системах для запроса отклика от удаленной станции. Этот символ может использоваться в качестве WRU¹ для получения информации или статуса удаленной станции.

ACK (Acknowledge) — подтверждение

Символ управления коммуникациями, передаваемый получателем в качестве отклика отправителю.

BEL (Bell) — звонок

Символ, используемый в тех случаях, когда требуется привлечь внимание человека. Он может управлять различными устройствами сигнализации.

BS (Backspace) — забой

Символ форматирования, обеспечивающий перевод печатающей головки на одну позицию назад в той же строке (символ применим и для устройств отображения).

HT (Horizontal Tabulation) — горизонтальная табуляция

Символ форматирования, обеспечивающий перевод печатающей головки в следующую позицию из группы предопределенных позиций в строке печати (символ применяется также для устройств отображения и в качестве средства пропуска перфокарт).

LF (Line Feed) — перевод строки

Символ форматирования, обеспечивающий перевод печатающей головки на следующую строку (символ применим и для устройств отображения). В тех случаях, когда это применимо, символ может трактоваться как NL² с переводом печатающей головки в начало следующей строки. Такое использование символа требует согласования между отправителем и получателем данных.

VT (Vertical Tabulation) — вертикальная табуляция

Символ форматирования, обеспечивающий перевод печатающей головки на следующую строку из группы предопределенных строк вывода (этот символ применим и для устройств отображения).

FF (Form Feed) — перевод страницы

Символ форматирования, обеспечивающий перевод печатающей головки на первую предопределенную строку следующей формы или страницы (этот символ применим и для устройств отображения).

CR (Carriage Return) — возврат каретки

Символ форматирования, обеспечивающий перевод печатающей головки в начало текущей строки (этот символ применим и для устройств отображения).

SO (Shift Out) — начало специальной трактовки символов

Управляющий символ, который показывает, что для следующей за ним последовательности символов, вплоть до Shift In, нужно использовать особую трактовку (не так, как символы стандартной таблицы кодов).

SI (Shift In) — возврат к стандартной трактовке символов

Управляющий символ, который указывает возврат к стандартной трактовке следующих за ним символов.

DLE (Data Link Escape) — особая трактовка символов

Символ управления коммуникациями, который меняет трактовку ограниченного числа следующих на нем непрерывно символов. Этот символ применяется исключительно для дополнительного управления в сетях передачи данных.

DC1, DC2, DC3, DC4 (Device Controls) — управление устройством

Символы, используемые для управления вспомогательными устройствами, связанными с обработкой или передачей данных, а именно, для включения (on) и выключения (off). Если требуется один сигнал «стоп» для прерывания работы или отключения вспомогательного устройства, предпочтительно использовать символ DC4.

NAK (Negative Acknowledge)

Символ управления коммуникациями, который принимающее устройство передает отправителю в качестве негативного подтверждения (отказа).

SYN (Synchronous Idle) — синхронизация при бездействии

Символ управления коммуникациями, используемый в синхронных системах передачи данных в отсутствии любых других символов (бездействие) для обеспечения синхронизации устройств связи.

ETB (End of Transmission Block) — завершение передаваемого блока

Символ управления коммуникациями, используемый для индикации завершения блока данных в коммуникациях. ETB используется в тех случаях, когда структура блока может быть не связана с форматом обработки данных.

CAN (Cancel) — отказ

Сигнал управления, показывающий, что данные, с которыми он был передан, отправлены по ошибке или должны игнорироваться.

¹Who Are You — буквально, «как вы?»

²New Line — новая строка.

EM (End of Medium) — конец среды (носителя)

Связанный с отправленными данными символ управления, который может использоваться для идентификации физического завершения носителя (среды), а также завершения используемой или запрошенной части записываемой на носитель информации (положение этого символа может не соответствовать физическому завершению носителя).

SUB (Substitute) — замена

Символ, который может использоваться для замены другого символа, признанного ошибочным или недействительным.

ESC (Escape)

Управляющий символ, предназначенный для расширения кодового набора (добавления символов) при обмене данными общего назначения. Символ Escape сам по себе служит префиксом, изменяющим интерпретацию ограниченного числа символов, непрерывно следующих за ним.

FS (File Separator), GS (Group Separator), RS (Record Separator), US (Unit Separator) - разделители

Эти символы могут служить разделителями информации с учетом иерархии — FS (наиболее включительный), GS, RS, US (содержимое и размер файла, группы, записи или блока не задаются).

DEL (Delete)

Этот символ служит для удаления (erase) или уничтожения (obliterate) ошибочных или нежелательных символов на перфоленте (строго говоря, символ DEL не является управляющим).

5.3 Символы графики

SP (Space)

Обычно не печатаемый графический символ, используемый для разделения слов. Служит также для форматирования вывода, смещая печатающую головку на одну позицию вперед (этот символ применим и для устройств отображения).

6. Общие вопросы

- 6.1 Этот стандарт не определяет средств, с помощью которых набор символов записывается на тот или иной физический носитель (среду), и не включает каких-либо методов резервирования и контроля ошибок. Более того, стандарт не определяет коммуникационных структур данных, форматов их представления, методов расширения кода или графического представления символов.
- 6.2 Отклонения от стандарта могут породить серьезные осложнения при обмене информацией, поэтому их следует допускать лишь с согласия всех вовлеченных сторон.
- 6.3 Относительный порядок в любой паре символов при их сортировке определяется двоичными значениями их кодов.
- 6.4 Не предусматривается никакой трактовки графических символов за исключением понятной пользователям. Кроме того, стандарт не задает стиля символов при печати или выводе на экран тех или иных графических символов. В конкретных приложениях может оказаться желательным использование разных стилей для отдельных графических символов при их использовании с конкретными целями. Например, символы с кодами 2/1 и 5/15 зачастую связывают с логическими операциями ИЛИ - OR (|) и НЕТ - NOT (252), соответственно.
- 6.5 Приложения к этому стандарту содержат дополнительную информацию об устройстве и применении кода .

[Этот документ RFC был переведен в машиночитаемую форму для включения в сетевые архивы RFC Робертом Беннетом в сентябре 1999 г.]

Перевод на русский язык

Николай Малых

nmalykh@gmail.com