

Википедия

Управляющие символы

Материал из Википедии – свободной энциклопедии

Управляющие символы — символы в кодировке, которым не приписано графическое представление, но которые используются для управления устройствами, организации передачи данных и других целей.

Сейчас для этих целей применяются форматы файлов, языки управления устройствами (такие как Postscript) и сетевые протоколы. Поэтому многие управляющие символы сейчас или не используются вообще, или используются не по назначению.

Стандарт POSIX требует обязательного наличия лишь восьми управляющих символов — \0, \a, \b, \t, \n, \v, \f, \r (см. переносимый набор символов).

Содержание
<u>Управляющие символы ASCII</u>
<u>Управляющие символы ISO 8859</u>
<u>Управляющие символы Unicode</u>
<u>Примечания</u>
<u>Ссылки</u>

Управляющие символы ASCII

В 1963 году была принята первая версия таблицы ASCII, в которой было определено 29 управляющих символов, большинство из которых находилось в диапазоне 00–1F, за исключением символов ACK,ESC и DEL. В редакции от 1967 года символы ACK и ESC были перенесены в основной диапазон, а их место заняли строчные буквы.

Управляющие символы ASCII предназначались для управления работой телетайпов и видеотерминалов и вводились на них сочетаниями с клавишей Ctrl, которая обнуляла в коде введённой клавиши бит 6. В современных компьютерных системах нигде, кроме эмуляторов терминала, не предусмотрен ввод этих символов напрямую (кроме символов табуляции и перевода строки) и большинство из перечисленных управляющих символов не используется.

Номер	Английское название	Русское название	Сочетание клавиш ^[1]	Escape последовательность	Назначение
<u>0x00</u>	NULL	пустой символ	^@	\0	Этот символ ничего не делает. Некоторые терминалы изображают его как пробел, но это неправильно. Часто NULL используют для обозначения конца цепочки символов (например, в языке C).
<u>0x01</u>	START OF HEADING	начало заголовка	^A		В настоящее время используется в консоли маршрутизаторов Cisco.
<u>0x02</u>	START OF TEXT	начало текста	^B		В настоящее время используется в консоли маршрутизаторов Cisco.
<u>0x03</u>	END OF TEXT	конец текста	^C		При вводе на терминале обычно интерпретируется как сигнал прерывания.
<u>0x04</u>	END OF TRANSMISSION	конец передачи	^D		При вводе на терминале в UNIX-системах интерпретируется как конец вводимых данных. Если текущая программа брала данные с терминала, то она завершается, как только обработает всё, что было до символа ^D.
<u>0x05</u>	ENQUIRY	запрос	^E		Использовался в телетайпной связи. В ответ предполагалось получить идентификационную строку удалённого аппарата.
<u>0x06</u>	ACKNOWLEDGE	подтверждение	^F		Использовался в телетайпной связи.
<u>0x07</u>	BELL	<u>звуковой сигнал</u>	^G	\a	Если этот символ послать на принтер или на терминал, то ничего не напечатается, но послышится звуковой сигнал.
<u>0x08</u>	BACKSPACE	возврат на шаг	^H	\b	Перемещает позицию печати на один символ назад. На принтерах может использоваться для наложения одного символа на другой, например a BS ^ = â . При вводе с терминала иногда используется для стирания предшествующего символа («забой»).
<u>0x09</u>	CHARACTER TABULATION (horizontal tabulation)	<u>горизонтальная табуляция</u>	^I	\t	Перемещает позицию печати к следующей позиции горизонтальной табуляции.
<u>0x0A</u>	LINE FEED	<u>перевод строки</u>	^J	\n	Перемещает позицию печати на одну строку вниз (исходно — без <i>возврата каретки</i>). Разделяет строки <u>текстовых файлов</u> в Unix-системах.
<u>0x0B</u>	LINE TABULATION (vertical tabulation)	вертикальная табуляция	^K	\v	Перемещает позицию печати к следующей позиции вертикальной табуляции. На терминалах этот символ обычно эквивалентен переводу строки.
<u>0x0C</u>	FORM FEED	прогон страницы, смена страницы	^L	\f	Выбрасывает текущую страницу и начинает печать со следующей. При выводе на терминал этот символ обычно эквивалентен переводу строки, однако нажатие ^L на клавиатуре может осуществлять очистку экрана.
<u>0x0D</u>	CARRIAGE RETURN	<u>Возврат каретки</u>	^M	\r	Перемещает позицию печати в крайнее левое положение (исходно — без <i>перевода на следующую строку</i>). Разделяет строки <u>текстовых файлов</u> в некоторых ОС (например Mac OS, но не в Mac OS X). Во многих других ОС (CP/M, MS-DOS и Microsoft Windows), для разделения строк используется сочетание кодов возврата каретки (CARRIAGE RETURN) и

					перевода строки (LINE FEED) — $0D_{16} + 0A_{16}$, то есть в том виде, в котором файл можно отправить непосредственно на принтер.
<u>0x0E</u>	SHIFT OUT (locking-shift one)	режим национальных символов	^N		В КОИ-7 включает режим национальных символов. На некоторых принтерах включает режим символов двойной ширины.
<u>0x0F</u>	SHIFT IN (locking-shift zero)	режим обычного ASCII	^O		В КОИ-7 включает латинский режим. На некоторых принтерах включает режим узких символов.
<u>0x10</u>	DATA LINK ESCAPE	экранирование управляющих символов (в обычный текст)	^P		Означает, что следующий за ним управляющий символ должен восприниматься как данные, а не как управляющий символ.
<u>0x11</u>	DEVICE CONTROL ONE	1-й код управления устройством	^Q		На терминалах разрешает продолжить вывод данных.
<u>0x12</u>	DEVICE CONTROL TWO	2-й код управления устройством	^R		
<u>0x13</u>	DEVICE CONTROL THREE	3-й код управления устройством	^S		На терминалах временно прерывает (приостанавливает) вывод данных.
<u>0x14</u>	DEVICE CONTROL FOUR	4-й код управления устройством	^T		
<u>0x15</u>	NEGATIVE ACKNOWLEDGE	отрицательное подтверждение	^U		Использовался в <u>телетайпной</u> связи.
<u>0x16</u>	SYNCHRONOUS IDLE	пустой символ для синхронного режима передачи	^V		Некоторые линии связи устроены так, что требуют непрерывной передачи данных. Если передавать нечего, то передают этот символ.
<u>0x17</u>	END OF TRANSMISSION BLOCK	конец блока передаваемых данных	^W		
<u>0x18</u>	CANCEL	отмена	^X		Данные, которые идут перед ним, некорректны. (Обычно речь идёт об одной строке.)
<u>0x19</u>	END OF MEDIUM	конец носителя	^Y		Использовался, например, если закончилась <u>перфолента</u> и т. п.
<u>0x1A</u>	SUBSTITUTE	символ замены	^Z		Ставится на месте символов, значения которых были потеряны при передаче. В <u>CP/M</u> и <u>MS-DOS</u> использовался для обозначения конца текстовых файлов и конца вводимых с консоли данных (хотя для этого были предназначены символы ^C и ^D).
<u>0x1B</u>	ESCAPE	Альтернативный регистр № 2 (AP2)	^[\textbackslash e	Означает, что следующие за ним символы имеют какое-то другое значение, отличное от того, которое определено в ASCII. Обычно начинает <u>управляющие последовательности</u> . См. также <u>ANSI.SYS</u> .
<u>0x1C</u>	INFORMATION SEPARATOR FOUR (file separator)	разделитель данных № 4 (разделитель файлов)	^\\		
<u>0x1D</u>	INFORMATION SEPARATOR	разделитель данных № 3	^]		

	THREE (group separator)	(разделитель групп)			
0x1E	INFORMATION SEPARATOR TWO (record separator)	разделитель данных № 2 (разделитель записей)	^^		Видимо, предназначался для разделения записей в <u>базах данных</u> , но практически никогда не используется для этого.
0x1F	INFORMATION SEPARATOR ONE (unit separator)	разделитель данных № 1 (разделитель полей)	^ —		Видимо, предназначался для разделения полей в базах данных, но практически никогда не используется для этого.
0x7F	DELETE	удаление	^?		Предназначен для забивания ошибочно пробитых символов на семидорожечных <u>перфолентах</u> (поскольку обозначается пробитием дырочек во всех дорожках), поэтому там он эквивалентен пустому символу (␣). На терминалах может генерироваться нажатием либо кнопки Backspace, либо кнопки Delete.

Управляющие символы ISO 8859

80, **PADDING CHARACTER**, символ-заполнитель.

81, **HIGH OCTET PRESET**, высокое октетное представление.

82, **BREAK PERMITTED HERE**, здесь разрешён разрыв строки.

83, **NO BREAK HERE**, здесь не разрешён разрыв строки.

84, **INDEX**, индекс.

85, **NEXT LINE**, следующая строка. Одновременно переводит строку и возвращает позицию печати к началу строки (эквивалентно `\r\n`).

86, **START OF SELECTED AREA**, начало выделенной области.

87, **END OF SELECTED AREA**, конец выделенной области.

88, **CHARACTER TABULATION SET**, установка позиций горизонтальной табуляции.

89, **CHARACTER TABULATION WITH JUSTIFICATION**, установка позиций и выравнивания горизонтальной табуляции.

8A, **LINE TABULATION SET**, установка позиций вертикальной табуляции.

8B, **PARTIAL LINE FORWARD**, частичный перевод строки вперёд.

8C, **PARTIAL LINE BACKWARD**, частичный перевод строки назад.

8D, **REVERSE LINE FEED**, обратный перевод строки.

8E, **SINGLE SHIFT TWO**, 2-е значение для следующего символа.

8F, **SINGLE SHIFT THREE**, 3-е значение для следующего символа.

90, **DEVICE CONTROL STRING**, строка управления устройством.

91, **PRIVATE USE ONE**, пользовательский символ № 1.

92, PRIVATE USE TWO, пользовательский символ № 2.

93, SET TRANSMIT STATE, установка режима передачи.

94, CANCEL CHARACTER, символ отмены.

95, MESSAGE WAITING, есть сообщение.

96, START OF GUARDED AREA, начало защищённой области.

97, END OF GUARDED AREA, конец защищённой области.

98, START OF STRING, начало строки.

99, SINGLE GRAPHIC CHARACTER INTRODUCER, следующий символ интерпретируется как специальный графический.

9A, SINGLE CHARACTER INTRODUCER, следующий символ интерпретируется как управляющий.

9B, CONTROL SEQUENCE INTRODUCER, начало управляющей последовательности. Обычно этот символ эквивалентен Escape+**[**.

9C, STRING TERMINATOR, окончание строки.

9D, OPERATING SYSTEM COMMAND, команда операционной системы.

9E, PRIVACY MESSAGE, секретное сообщение.

9F, APPLICATION PROGRAM COMMAND, команда прикладной программы.

Управляющие символы Unicode

034F, COMBINING GRAPHEME JOINER. Объединить символы, стоящие слева и справа (создать лигатуру).

2008, PUNCTUATION SPACE, пробел шириной в пунктуацию.

200B, ZERO-WIDTH SPACE, пробел нулевой ширины. При выравнивании по ширине может расширяться.

200C, ZERO WIDTH NON-JOINER. Запрещает образование лигатур.

200D, ZERO WIDTH JOINER. Разрешает образование лигатур.

200E, LEFT-TO-RIGHT MARK. Писать слева направо.

200F, RIGHT-TO-LEFT MARK. Писать справа налево.

2028, LINE SEPARATOR, разделитель строк. Разделяет строки текста, но не абзацы.

2029, PARAGRAPH SEPARATOR, разделитель абзацев. Разделяет абзацы текста.

202A, LEFT-TO-RIGHT EMBEDDING. Начало текста, написанного слева направо, внутри текста, написанного справа налево.

202B, RIGHT-TO-LEFT EMBEDDING. Начало текста, написанного справа налево, внутри текста, написанного слева направо.

202C, POP DIRECTIONAL FORMATTING. Конец вставленного текста с другим направлением.

202D, LEFT-TO-RIGHT OVERRIDE. Заменить текст, написанный слева направо, текстом, написанным справа налево.

202E, RIGHT-TO-LEFT OVERRIDE. Заменить текст, написанный справа налево, текстом, написанным слева направо.

2060, WORD JOINER, соединитель слов.

2061, FUNCTION APPLICATION, применение функции.

2063, INVISIBLE SEPARATOR, невидимый разделитель.

2066, LEFT-TO-RIGHT ISOLATE, разделитель слева-направо.

2067, RIGHT-TO-LEFT ISOLATE, разделитель справа-налево.

2068, FIRST STRONG ISOLATE, первый усиленный разделитель.

2069, POP DIRECTIONAL ISOLATE, разделитель популярного направления.

206A, INHIBIT SYMMETRIC SWAPPING, запретить симметричный обмен^[2].

206B, ACTIVATE SYMMETRIC SWAPPING, активировать симметричный обмен^[2].

206C, INHIBIT ARABIC FORM SHAPING, запретить арабскую форму формирования^[2].

206D, ACTIVATE ARABIC FORM SHAPING, активировать арабскую форму формирования^[2].

206E, NATIONAL DIGIT SHAPES, национальная форма цифры^[2].

206F, NOMINAL DIGIT SHAPES, номинальная форма цифры^[2].

FE00 ... FE0F, VARIATION SELECTOR-1...16, выбор варианта начертания № 1 ... № 16.

FEFF, ZERO WIDTH NO-BREAK SPACE / BYTE ORDER MARK, неразрывный пробел нулевой ширины / индикатор порядка байтов. Этот символ используется для указания того, что данный файл записан в UTF-16 или UTF-32 с определённым порядком байтов (поскольку символа FFFE нет, а в UTF-8 байты FE и FF не используются). Использование этого символа в качестве неразрывного пробела нулевой ширины, несмотря на первоначальное название «ZERO WIDTH NO-BREAK SPACE», не рекомендуется; для этого есть символ U+2060 (word joiner).

FFF9, INTERLINEAR ANNOTATION ANCOR, якорь межстрочной аннотации.

FFFA, INTERLINEAR ANNOTATION SEPARATOR, разделитель межстрочной аннотации.

FFFB, INTERLINEAR ANNOTATION TERMINATOR, терминатор межстрочной аннотации.

FFFC, OBJECT REPLACEMENT CHARACTER, объектно-заменяющий символ.

FFFD, REPLACEMENT CHARACTER, заменяющий символ. Используется, когда значение символа неизвестно или не может быть выражено в Юникоде (см. также символ 1A).

E0100 ... E01EF, VARIATION SELECTOR-17...256, выбор варианта начертания № 17 ... № 256.

Примечания

1. Для первых 32 символов это комбинация Ctrl и клавиши с символом с кодом на 64 больше, чем код управляющего символа.
2. Символ не рекомендуется к использованию.

Ссылки

- ISO IR 1 (<https://www.itscj.ipsj.or.jp/iso-ir/001.pdf>) — C0 Set of ISO 646 (PDF)

Источник — https://ru.wikipedia.org/w/index.php?title=Управляющие_символы&oldid=123057757

Эта страница в последний раз была отредактирована 5 июня 2022 в 10:03.

Текст доступен по лицензии Creative Commons Attribution-ShareAlike; в отдельных случаях могут действовать дополнительные условия.

Wikipedia® — зарегистрированный товарный знак некоммерческой организации Wikimedia Foundation, Inc.