

Chapter IV-7 — Programming Techniques

Igor Pattern	PCRE Pattern	Character Matched
\\t	\t	Tab (hex 09)
\\ddd	\ddd	Character with octal code ddd, or backreference
\\xhh	\xhh	Character with hex code hh

Backslash and Nonprinting Characters Arcania

The material in this section is arcane and rarely needed. We recommend that you skip it.

The precise effect of `\cx` is if `x` is a lower case letter, it is converted to upper case. Then bit 6 of the character (hex 40) is inverted. Thus `\cz` becomes hex 1A, but `\c{` becomes hex 3B, while `\c;` becomes hex 7B.

After `\x`, from zero to two hexadecimal digits are read (letters can be in upper or lower case). If characters other than hexadecimal digits appear between `\x{` and `}`, or if there is no terminating `}`, this form of escape is not recognized. Instead, the initial `\x` will be interpreted as a basic hexadecimal escape, with no following digits, giving a character whose value is zero.

After `\0` up to two further octal digits are read. In both cases, if there are fewer than two digits, just those that are present are used. Thus the sequence `\0\x\07` specifies two binary zeros followed by a BEL character (code value 7). Make sure you supply two digits after the initial zero if the pattern character that follows is itself an octal digit.

The handling of a backslash followed by a digit other than 0 is complicated. Outside a character class, PCRE reads it and any following digits as a decimal number. If the number is less than 10, or if there have been at least that many previous capturing left parentheses in the expression, the entire sequence is taken as a back reference. A description of how this works is given later, following the discussion of parenthesized subpatterns.

Inside a character class, or if the decimal number is greater than 9 and there have not been that many capturing subpatterns, PCRE rereads up to three octal digits following the backslash, and generates a single byte from the least significant 8 bits of the value. Any subsequent digits stand for themselves. For example:

Igor Pattern	PCRE Pattern	Character(s) Matched
\\040	\040	Another way of writing a space
\\40	\40	A space, provided there are fewer than 40 previous capturing subpatterns
\\7	\7	Always a back reference
\\11	\11	Might be a back reference, or another way of writing a tab
\\011	\011	Always a tab
\\0113	\0113	A tab followed by the character "3"
\\113	\113	Might be a back reference, otherwise the character with octal code 113
\\377	\377	Might be a back reference, otherwise the byte consisting entirely of 1 bits
\\81	\81	Either a back reference or a binary zero followed by the two characters "8" and "1"

Note that octal values of 100 or greater must not be introduced by a leading zero, because no more than three octal digits are ever read.

All the sequences that define a single byte value can be used both inside and outside character classes. In addition, inside a character class, the sequence `\b` is interpreted as the backspace character (hex 08), and the sequence `\x` is interpreted as the character X. Outside a character class, these sequences have different meanings (see **Backslash and Nonprinting Characters** on page IV-179).