

Chapter IV-7 — Programming Techniques

Note: Igor doesn't use the Perl opening and closing regular expression delimiter character which is the forward slash. In Perl you would use "/Fred/" and "/ (?i) fred/".

Regular Expression Metacharacters

The power of regular expressions comes from the ability to include alternatives and repetitions in the pattern. These are encoded in the pattern by the use of "metacharacters", which do not stand for themselves but instead are interpreted in some special way.

There are two different sets of metacharacters: those that are recognized anywhere in the pattern except within brackets, and those that are recognized in brackets. Outside brackets, the metacharacters are as follows:

\	General escape character with several uses
^	Match start of string
\$	Match end of string
.	Match any character except newline (by default) (To match newline, see Matching Newlines on page IV-184)
[Start character class definition (for matching one of a set of characters)
	Start of alternative branch (for matching one or the other of two patterns)
(Start subpattern
)	End subpattern
?	0 or 1 quantifier (for matching 0 or 1 occurrence of a pattern) Also extends the meaning of (
	Also quantifier minimizer
*	0 or more quantifier (for matching 0 or more occurrence of a pattern)
+	1 or more quantifier (for matching 1 or more occurrence of a pattern) Also possessive quantifier
{	Start min/max quantifier (for matching a number or range of occurrences)

Character Classes in Regular Expressions

A character class is a set of characters and is used to specify that one character of the set should be matched. Character classes are introduced by a left-bracket and terminated by a right-bracket. For example:

[abc]	Matches a or b or c
[A-Z]	Matches any character from A to Z
[A-Za-z]	Matches any character from A to Z or a to z.

POSIX character classes specify the characters to be matched symbolically. For example:

[:alpha:]	Matches any alphabetic character (A to Z or a to z).
[:digit:]	Matches 0 to 9.

In a character class the only metacharacters are:

\	General escape character
^	Negate the class, but only if ^ is the first character
-	Indicates character range
[POSIX character class (only if followed by POSIX syntax)