

Chapter IV-7 — Programming Techniques

Circumflex need not be the first character of the pattern if a number of alternatives are involved, but it should be the first thing in each alternative in which it appears if the pattern is ever to match that branch. If all possible alternatives start with a circumflex, that is, if the pattern is constrained to match only at the start of the subject, it is said to be an “anchored” pattern. (There are also other constructs that can cause a pattern to be anchored.)

A dollar character \$ is an assertion that is true only if the current matching point is at the end of the subject string, or immediately before a newline character that is the last character in the string (by default). Dollar need not be the last character of the pattern if a number of alternatives are involved, but it should be the last item in any branch in which it appears. Dollar has no special meaning in a character class.

Dot, Period, or Full Stop

Outside a character class, a dot in the pattern matches any one character in the subject, including a nonprinting character, but not (by default) newline. Dot has no special meaning in a character class.

The match option setting (?s) changes the default behavior of dot so that it matches any character including newline. The setting can appear anywhere before the dot in the pattern. See **Matching Newlines** on page IV-184 for details.

Character Classes and Brackets

An opening bracket introduces a character class which is terminated by a closing bracket. A closing bracket on its own is not special. If a closing bracket is required as a member of the class, it must be the first data character in the class (after an initial circumflex, if present) or escaped with a backslash.

A character class matches a single character in the subject. A matched character must be in the set of characters defined by the class, unless the first character in the class definition is a circumflex, in which case the subject character must not be in the set defined by the class. If a circumflex is actually required as a member of the class, ensure it is not the first character, or escape it with a backslash.

For example, the character class [aeiou] matches any English lower case vowel, whereas [^aeiou] matches any character that is not an English lower case vowel. Note that a circumflex is just a convenient notation for specifying the characters that are in the class by enumerating those that are not.

When caseless matching is set, using the (?i) match option setting, any letters in a class represent both their upper case and lower case versions, so for example, the caseless pattern (?i) [aeiou] matches A as well as a, and the caseless pattern (?i) [^aeiou] does not match A.

The minus (hyphen) character can be used to specify a range of characters in a character class. For example, [d-m] matches any letter between d and m, inclusive. If a minus character is required in a class, it must be escaped with a backslash or appear in a position where it cannot be interpreted as indicating a range, typically as the first or last character in the class.

To include a right-bracket in a range you must use \]. As usual, this would be represented in a literal Igor string as \\].

Though it is rarely needed, you can specify a range using octal numbers, for example [\000-\037].

The character types \d, \D, \p, \P, \s, \S, \w, and \W may also appear in a character class, and add the characters that they match to the class. For example, [\dABCDEF] matches any hexadecimal digit. A circumflex can conveniently be used with the upper case character types to specify a more restricted set of characters than the matching lower case type. For example, the class [^\W_] matches any letter or digit, but not underscore. The corresponding Grep command would begin with

```
Grep/E="[^\\W_]"...
```

The only metacharacters that are recognized in character classes are backslash, hyphen (only where it can be interpreted as specifying a range), circumflex (only at the start), opening bracket (only when it can be interpreted as introducing a POSIX class name — see **POSIX Character Classes** on page IV-183), and the terminating closing bracket. However, escaping other nonalphanumeric characters does no harm.