

## Chapter IV-7 — Programming Techniques

However, if a quantifier is followed by a question mark, it ceases to be greedy, and instead matches the minimum number of times possible, so the pattern

```
/\.*.*?\*/      or      Grep/E="/\\.*.*?\\*/"
```

does the right thing with the C comments. The meaning of the various quantifiers is not otherwise changed, just the preferred number of matches.

Do not confuse this use of question mark with its use as a quantifier in its own right. Because it has two uses, it can sometimes appear doubled, as in

```
\d??\d      or      Grep/E="\d??\\d"
```

which matches one digit by preference, but can match two if that is the only way the rest of the pattern matches.

If the PCRE\_UNGREEDY option (?U) is set, the quantifiers are not greedy by default, but individual ones can be made greedy by following them with a question mark. In other words, it inverts the default behavior.

### Quantifiers With Subpatterns

When a parenthesized subpattern is quantified with a minimum repeat count that is greater than 1 or with a limited maximum, more memory is required for the compiled pattern, in proportion to the size of the minimum or maximum.

When a capturing subpattern is repeated, the value captured is the substring that matched the final iteration. For example, after

```
(tweedle[dume]{3}\s*)+      or      Grep/E="(tweedle[dume]{3}\\s*)+"
```

has matched “tweedledum tweedledee” the value of the captured substring is “tweedledee”. However, if there are nested capturing subpatterns, the corresponding captured values may have been set in previous iterations. For example, after

```
/ (a | (b) ) +/
```

matches “aba” the value of the second captured substring is “b”.

### Atomic Grouping and Possessive Quantifiers

With both maximizing and minimizing repetition, failure of what follows normally reevaluates the repeated item to see if a different number of repeats allows the rest of the pattern to match. Sometimes it is useful to prevent this, either to change the nature of the match, or to cause it fail earlier than it otherwise might, when the author of the pattern knows there is no point in carrying on.

Consider, for example, the pattern \d+foo when applied to the subject line

```
123456bar
```

After matching all 6 digits and then failing to match “foo”, the normal action of the matcher is to try again with only 5 digits matching the \d+ item, and then with 4, and so on, before ultimately failing. “Atomic grouping” provides the means for specifying that once a subpattern has matched, it is not to be reevaluated in this way.

If we use atomic grouping for the previous example, the matcher would give up immediately on failing to match “foo” the first time. The notation is a kind of special parenthesis, starting with (?> as in this example:

```
(?>\d+) foo      or      Grep/E="(?>\\d+) foo"
```

This kind of parenthesis “locks up” the part of the pattern it contains once it has matched, and a failure further into the pattern is prevented from backtracking into it. Backtracking past it to previous items, however, works as normal.

An alternative description is that a subpattern of this type matches the string of characters that an identical standalone pattern would match, if anchored at the current point in the subject string.

Atomic grouping subpatterns are not capturing subpatterns. Simple cases such as the above example can be thought of as a maximizing repeat that must swallow everything it can. So, while both \d+ and \d+?