This is a quick reference to Perl's regular expressions. For full information see the perlre and perlop manual pages.

Operators

e = " determines to which variable the regex is applied. In its absence, $ is used.

! = " determines to which variable the regex is applied, and negates the result of the match; it returns false if the match succeeds, and true if it fails.

m/search/ 

searches a string for a pattern match, applying the given options.

i case-insensitive

m multiline mode -- " and $ match internal lines

s match as a single line -- matches 

o compile pattern once

x extended legibility -- free whitespace and comments

c don't reset pos on failed matches when using /g

If pattern is an empty string, the last successfully matched regex is used. Delimiters other than '//' may be used for both this operator and the following ones.

qr/search/ 

lets you store a regex in a variable, or pass one around. Modifiers as for m/ and are stored within the regex.

s/search/replace/ 

substitutes matches of pattern with replacement. Modifiers as for m/ with one addition:

e evaluate replacement as an expression

' e ' may be specified multiple times. replacement is interpreted as a double quoted string unless a single-quote (') is the delimiter.

?pattern? is like m/pattern/ but matches only once. No alternate delimiters can be used. Must be reset with \reset.

Syntax

\ escapes the character immediately following it
.
Matches any single character except a newline (unless /s is used)

Matches at the beginning of the string (or line, if /m is used)

Matches at the end of the string (or line, if /m is used)
+ Matches the preceding element 1 or more times

? Matches the preceding element 0 or 1 times

{} Specifies a range of occurrences for the element preceding it

[... ] Matches any one of the characters contained within the brackets

( ) Groups subexpressions for capturing to $1, $2...

{...,} Groups subexpressions without capturing (cluster)

| Matches either the subexpression preceding or following it

\1, \2 ... The text from the Nth group

Escape sequences

These work as in normal strings.

\a Alarm (beep)

\b Backspace

\c Control-x

\d A digit

\D A nondigit

\w A word character

\W A non-word character

\s A whitespace character

\S A non-whitespace character

\C Match a byte (with Unicode, ‘.’ matches a character)

\P Match P-named (Unicode) property

\p{...} Match lack of Unicode property with long name

\p{...} Match Unicode property with long name

\X Match extended unicode sequence

POSIX character classes and their Unicode and Perl equivalents:

<table>
<thead>
<tr>
<th>POSIX</th>
<th>Unicode</th>
<th>Perl</th>
</tr>
</thead>
<tbody>
<tr>
<td>alnum</td>
<td>IsAlnum</td>
<td></td>
</tr>
<tr>
<td>alpha</td>
<td>IsAlpha</td>
<td></td>
</tr>
<tr>
<td>ascii</td>
<td>IsASCII</td>
<td></td>
</tr>
<tr>
<td>blank</td>
<td>IsSpace</td>
<td></td>
</tr>
<tr>
<td>cntrl</td>
<td>IsCtrl</td>
<td></td>
</tr>
<tr>
<td>digit</td>
<td>IsDigit</td>
<td></td>
</tr>
<tr>
<td>graph</td>
<td>IsGraph</td>
<td></td>
</tr>
<tr>
<td>lower</td>
<td>IsLower</td>
<td></td>
</tr>
<tr>
<td>print</td>
<td>IsPrint</td>
<td></td>
</tr>
<tr>
<td>punct</td>
<td>IsPunct</td>
<td></td>
</tr>
<tr>
<td>space</td>
<td>IsSpace</td>
<td></td>
</tr>
<tr>
<td>upper</td>
<td>IsUpper</td>
<td></td>
</tr>
<tr>
<td>word</td>
<td>IsWord</td>
<td></td>
</tr>
<tr>
<td>xdigit</td>
<td>IsXDigit</td>
<td></td>
</tr>
</tbody>
</table>

Within a character class:

POSIX traditional Unicode

\[digit:] \ [P{IsDigit}]

\[^digit:] \ [P{!IsDigit}]

Anchors

All are zero-width assertions.

- Match string start (or line, if /m is used)

\$ Match string end (or line, if /m is used) or before newline

\b Match word boundary (between \w and \W)

\B Match except at word boundary (between \w or \W)

\A Match string start (regardless of /m)

\Z Match string end (before optional newline)

\G Match absolute string end

\G Match where previous m/g left off
Quantifiers
Quantifiers are greedy by default – match the longest leftmost.

Maximal          Minimal          Allowed range
(n,m)?           (n,m)?           Must occur at least n times but no more than m times
{n,}?            {n,}?            Must occur at least n times
{n}?             {n}?             Must occur exactly n times
*                 +?               0 or more times (same as {0,})
+                 +?               1 or more times (same as {1,})
?                 ??               0 or 1 time (same as {0,1})

There is no quantifier {,n} – that gets understood as a literal string.

Extended constructs
(?#text)          A comment
(?!msx-!msx:...)  Enable/disable option (as per m// modifiers)
(?:...)           Zero-width positive lookahead assertion
(?!...)           Zero-width negative lookahead assertion
(?<...)           Zero-width positive lookbehind assertion
(?<...)           Zero-width negative lookbehind assertion
(?{ code })      Grab what we can, prohibit backtracking
(?{ code })      Embedded code, return value becomes $^R
(?{ code })      Dynamic regex, return value used as regex
(?{cond}yes|no)   cond being integer corresponding to capturing parens
(?{cond}yes)     or a lookahead/eval zero-width assertion

Variables
$_            Default variable for operators to use
$*            Enable multiline matching (deprecated; not in 5.9.0 or later)
$&            Entire matched string
$'            Everything prior to matched string
$+            Everything after to matched string

The use of those last three will slow down all regex use within your program. Consult the perlvar man page for a description of how to use these variables to emulate Perl's own behaviour. Consult the perlvar man page for @LAST_MATCH_START and @LAST_MATCH_END to see equivalent expressions that won’t cause slow down. See also Devel::SawAmpersand.$1, $2...   Hold the Xth captured expr
$+            Last parenthesized pattern match
$'N           Holds the most recently closed capture
$'R           Holds the result of the last (?{...}) expr
$-            Offsets of starts of groups. $-[0] holds start of whole match
$+            Offsets of ends of groups. $+[0] holds end of whole match

Captured groups are numbered according to their opening paren.

Functions
lc            Lowercase a string
lcfirst       Lowercase first char of a string
uc            Uppercase a string
ucfirst       Titlecase first char of a string
pos            Return or set current match position
quotemeta     Quote metacharacters
reset          Reset ?pattern? status
study          Analyze string for optimizing matching
split          Use regex to split a string into parts

The first four of these are like the escape sequences \L, \l, \U, and \u. For Titlecase, see below.

Terminology
Titlecase
Unicode concept which most often is equal to uppercase, but for certain characters like the German 'sharp s' (&) there is a difference.

See also
• perlreut for a tutorial on regular expressions.
• perlreqn for a rapid tutorial.
• perlre for more details.
• perlde for details on the variables.
• perl for details on the operators.
• perlfun for details on the functions.
• perlfaq6 for FAQs on regular expressions.
• perlregimpl, perlunicode, charnames and locale for details on regexes and internationalisation.
• Mastering Regular Expressions by Jeffrey Friedl (http://regex.info/) for a thorough grounding and reference on the topic.

Authors
This card was created by Andrew Ford.

The original document (perlreref.pod) is part of the standard Perl distribution. It was written by Iain Truskett, with thanks to David P.C. Wollmann, Richard Soderberg, Sean M. Burke, Tom Christiansen, Jim Cromie, and Jeffrey Goff for useful advice.