Perl Regular Expression

Specific characters:

\t	A tab character
\n	A newline character (OS neutral)
\r	A carriage return character
\f	A form feed character
\cX	Control character **CTRL-X**
\NNN	Octal code for character **NNN**

Metacharacters:

The following 14 characters need to be escaped with a backslash - "****" - because by default, they mean something special.

\ | ( ) [ ] { } ^ $ * + ? .

.	Match any one character (except **\n)**
|	Alternation
( )	Group and capture
[ ]	Define character class
****	overrides the meaning of the next char.
{?.. }	specifies a non-capturing group

Anchors:

^	Match at the beginning of a string (or line)
$	Match at the end of a string (or line)
\b	Match at a 'word' boundary
\B	Match at not a 'word' boundary

These are also known as *zero width assertions.*

Extracting matches for ( )

$1, $2, .	the part that matched inside goes into the special variables. It can be used outside of regex itself.
\g1, \g2, …	matching variables that can be used inside a regex

Quantifiers:

These quantifiers apply to the preceding *atom.*

*****	Match 0 or more times
+	Match 1 or more times
?	Match 0 or 1 times
{N}	Match exactly ***N*** times
{N, }	Match at least ***N*** times
{N,M}	Match at least ***N*** but not more than ***M*** times

By default, quantifiers are *"greedy"*. They attempt to match as many characters as possible. In order to make them match as few characters as possible, follow them with a question mark "?".

Character class metacharacters:

^	If the first character of a class, negates that class
-	Unless first or last character of a class, used for a range

Character class shortcuts:

\d	**[0-9]**	A digit
\D	**[^0-9]**	A non-digit
\s	**[\t\n\r\f]**	A whitespace char.
\S	**[^\t\n\r\f]**	A non-whitespace char.
\w	**[a-zA-Z0-9]**	A 'word' char.
\w	**[^ a-zA-Z0-9]**	A 'non-word' char.

These shortcuts can be used either on their own, or within a character class.

Search and replace:

s/regex/replacement/modifiers

Metaquote & case translations:

\Q	**Quote** (de-meta) characters until **\E**
\U	Uppercase characters until **\E**
\L	Lowercase characters until **\E**

Special variables:

$ '	The characters to the left of the match
$ &	The characters that matched
$ '	The characters to the right of the match
\N	The characters captured by the Nth set of parentheses (if on the match side)
$N	The characters captured by the Nth set of parentheses (if not on the match side)

Modifiers:

These modifiers apply to the entire pattern

/i	Ignore case
/g	Match globally (all)
/m	Let **^** and $ match next to embedded **\n**
/s	Let **.** match **\n**
/x	Ignore most whitespace and allow comments
/e	Evaluate right hand side of s/ // as an expression

All except **/e** apply to both **m//** and **s///**.

Binding operators:

=~	True if the regex matches
!~	True if the regex doesn't match

Reference:

http://perldoc.perl.org/index-tutorials.html
http://perldoc.perl.org/perlrequick.html
http://perldoc.perl.org/perlretut.html