Regular expressions 1. Special characters

The following characters are the *meta* characters that give special meaning to the regular expression search syntax:

\ the backslash *escape* character.
The backslash gives special meaning to the character following it. For example, the combination "\n" stands for the *newline*, one of the control characters. The combination "\w" stands for a "word" character, one of the convenience escape sequences while "\1" is one of the substitution special characters.
Example: The regex "aa\n" tries to match two consecutive "a"s at the end of a line, inclusive the newline character itself.
Example: "a\+" matches "a+" and not a series of one or "a"s.

^ the caret is the start of line anchor or the negate symbol.
Example: "^a" matches "a" at the start of a line.
Example: "[^0-9]" matches any non digit.

$ the dollar is the end of line anchor.
Example: "b$" matches a "b" at the end of a line.
Example: "^b$" matches the empty line.

{ } the open and close curly bracket are used as range quantifiers.
Example: "a{2,3}" matches "aa" or "aaa".

[ ] the open and close square bracket define a character class to match a *single* character.
The "^" as the first character following the "[" negates and the match is for the characters *not* listed. The "-" denotes a range of characters. Inside a "[ ]" character class construction most special characters are interpreted as ordinary characters.
Example: "[d-f]" is the same as "[def]" and matches "d", "e" or "f".
Example: "[a-z]" matches any lowercase characters in the alfabet.
Example: "[^0-9]" matches any character that is not a digit.
Example: A search for "[][()?<>.*?]" in the string "[]()?<>.*?" followed by a replace string "r" has the result "rrrrrrrrrrrr". Here the search string is *one* character class and all the meta characters are interpreted as ordinary characters without the need to escape them.

( ) the open and close parenthesis are used for grouping characters (or other regex).
The groups can be referenced in both the search and the substitution phase. There also exist some special constructs with parenthesis.
Example: "(ab)\1" matches "abab".

. the dot matches any character except the newline.
Example: ".a" matches two consecutive characters where the last one is "a".
Example: ".*\.txt$" matches all strings that end in ".txt".

***** the star is the match-zero-or-more quantifier.
Example: "^.*$" matches an entire line.

+ the plus is the match-one-or-more quantifier.

? the question mark is the match-zero-or-one quantifier. The question mark is also used in special constructs with parenthesis and in changing match behaviour.

| the vertical pipe separates a series of alternatives.
Example: "(a|b|c)a" matches "aa" or "ba" or "ca".

< > the smaller and greater signs are anchors that specify a left or right word boundary.

- the minus indicates a range in a character class (when it is not at the first position after the "[" opening bracket or the last position before the "]" closing bracket.
Example: "[A-Z]" matches any uppercase character.
Example: "[A-Z-]" or "[-A-Z]" match any uppercase character or "-".

& the and is the "substitute complete match" symbol.

Links to this page

- Regular expressions
