

Metacharacters

- **[] → Square brackets**
 - Square brackets specifies a set of characters you wish to match.

Expression	String	Matched?
[abc]	a	1 match
	ac	2 matches
	Hey Jude	No match
	abc de ca	5 matches

Here, [abc] will match if the string you are trying to match contains any of the a, b or c.

You can also specify a range of characters using - inside square brackets.

- [a-e] is the same as [abcde].
- [1-4] is the same as [1234].

Metacharacters

- You can complement (invert) the character set by using caret ^ symbol at the start of a square-bracket.
 - [^abc] means any character except a or b or c.
 - [^0-9] means any non-digit character.

Metacharacters

- **. → Period**
 - A period matches any single character (except newline '\n').

Expression	String	Matched?
..	a	No match
	ac	1 match
	acd	1 match
	acde	2 matches (contains 4 characters)

Metacharacters

- **^ → Caret**
 - The caret symbol ^ is used to check if a string starts with a certain character.

Expression	String	Matched?
^a	a	1 match
	abc	1 match
	bac	No match
^ab	abc	1 match
	acb	No match (starts with a but not followed by b)

Metacharacters

- **\$ → Dollar**
 - The dollar symbol is used to check if a string ends with a certain character.

Expression	String	Matched?
a\$	a	1 match
	formula	1 match
	cab	No match

Metacharacters

- * → Star

- The star symbol matches zero or more occurrences of the pattern left to it.

Expression	String	Matched?
ma*n	mn	1 match
	man	1 match
	maan	1 match
	main	No match (a is not followed by n)
	woman	1 match

Metacharacters

- **+ → Plus**
 - The plus symbol matches one or more occurrences of the pattern left to it.

Expression	String	Matched?
ma+n	mn	No match (no a character)
	man	1 match
	maan	1 match
	main	No match (a is not followed by n)
	woman	1 match

Metacharacters

- **? → Question Mark**
 - The question mark symbol matches zero or one occurrence of the pattern left to it.

Expression	String	Matched?
ma?n	mn	1 match
	man	1 match
	maan	No match (more than one <code>a</code> character)
	main	No match (a is not followed by n)
	woman	1 match

Metacharacters

- **{}** → Braces

- Consider this code: `{n,m}`. This means at least `n`, and at most `m` repetitions of the pattern left to it.

Expression	String	Matched?
<code>a{2,3}</code>	<code>abc dat</code>	No match
	<code>abc daat</code>	1 match (at <code>daat</code>)
	<code>aabc daaat</code>	2 matches (at <code>aabc</code> and <code>daaat</code>)
	<code>aabc daaaat</code>	2 matches (at <code>aabc</code> and <code>daaaat</code>)

Metacharacters

- This RegEx `[0-9]{2,4}` matches at least 2 digits but not more than 4 digits.

Expression	String	Matched?
	ab123csde	1 match (match at ab <u>123</u> csde)
<code>[0-9]{2,4}</code>	12 and 345673	2 matches (at <u>12</u> and <u>345673</u>)
	1 and 2	No match

Metacharacters

- | → **Alternation:**
 - Vertical bar is used for alternation (or operator).

Expression	String	Matched?
a b	cde	No match
	ade	1 match (match at <u>a</u> de)
	acdbea	3 matches (at <u>a</u> <u>c</u> <u>d</u> bea)

Here, a|b match any string that contains either a or b

Metacharacters

- **() → Group**
 - Parentheses is used to group sub-patterns.
 - For example, (a|b|c)xz match any string that matches either a or b or c followed by xz.

Metacharacters

- **\ → Backslash**
 - Backslash \ is used to escape various characters including all metacharacters.
 - For example, \\$a match if a string contains \$ followed by a. Here, \$ is not interpreted by a RegEx engine in a special way.
 - If you are unsure if a character has special meaning or not, you can put \ in front of it. This makes sure the character is not treated in a special way.

Metacharacters

- A character class matches any one of a set of characters. It is used to match the most basic element of a language like a letter, a digit, space, a symbol etc.

/s : matches any whitespace characters such as space and tab

/S : matches any non-whitespace characters

/d : matches any digit character

/D : matches any non-digit characters

/w : matches any word character (basically alpha-numeric)

/W : matches any non-word character

- In compute programming, **whitespace is any character or series of characters that represent horizontal or vertical space in typography.**