Regular Expressions

Repetitive patterns in a text string

Starting with "a"

next at least one "b" or more

ab

abb

abbbbbbbbbbbbbbbb

abbbbbbbbbbbbabbbbbbbbbbbb

(ab+)+

Tokens

\w - Word [alphanumeric] (A-Z) or (a-z) or (0-9)

\W - Not word (negation of \w)

\s - Blank space " "

\S - Not blank space (not \s)

\d - Digit 0-9

\D - Not digit (not \d)

\t - Tab " "

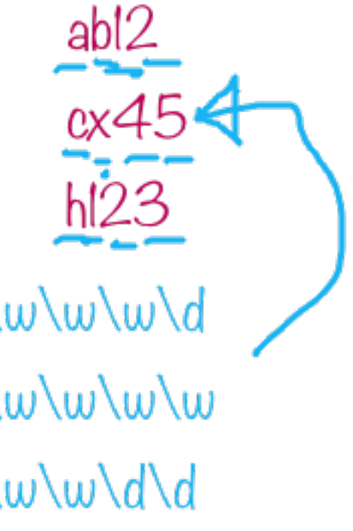
\T - Not tab (not \t)

\n - New line "\n" (line break)

\N - Not \n

. - Any character @ % $ & / ! ° - O y (Any)

\. - Point character hello.a (the point in a string)



Repeaters

u? - "u" zero ore one (optional)

u* - Repeat "u" zero or more times

u+ - Repeat "u" one or more times

u{3} - Repeat "u" exactly 3 times

u{3, 6} - Repeat "u" from 3 to 6 times

u{3, } - Repeat "u" 3 or more times

Pattern: (a-?)+

Matches: "a" "a-" "a-a" "a-a-" "a-aa-" "aa-aaa-aaa"

Pattern (a-*)+

Matches: "aaaa————aaaaaa"

Pattern: a*

Matches: "" "a" "aa" "aaa" "aaa...a"

Pattern: (aba)+

Matches: "aba" "abaaba" "abaabaaba" "abaabaaba...aba"

Pattern: (\w\d){5}

Matches: "a1b2c3d4e5" "1234567896" "X9X9X9X9_9"

Groups

(·) - Subsequence

[·] - Special groups

Special groups

[A-Z] - All characters in uppercase

[a-z] - All characters in lowercase

[A-Za-z] - All characters in uppercase or lowercase

[\s\t\n] - A blank space or tab or line break

[aeiou] - Lowercase vowels

[AaeEiIOouU] - All characters "A", "E", "I", "O", "U", "a", "e", "i", "o", "u"

"u" belongs to special group

String: 1000101110101101010101

Size: 22 chars

Pattern: [01]{22}

Pattern: (01) different !!! take care

A special group [.] is a set of valid tokens

Initializers and Endings

Strings starting with "hel"

Pattern: ^hel

Strings ending with "ld"

Pattern: ld$

Assembled two or more patterns

Pattern P
Pattern Q

We want to match P or Q

Pattern: P|Q

Pattern: P|Q|R|S|...

o_obadillo@hotmail.

Recognize Emails

Pattern: [\w\.-]+@[\w]{3, }\.([\w\.]{2, })+