\d = atalho para dígitos 0-9

A classe \D, por sua vez, é um non-digit, ou seja, \D é um atalho para [^\d]

\s = espaços em branco whitespace \*OBS1\* Experimente usar \s+

[xxx] = COLCHETES para definir classes de caracteres

\w significa word char e é um atalho para [A-Za-z0-9\_].

O \W é a non-word char, ou seja tudo que não é um word char. \W é um atalho para [^\w].

back-references, uma referência para um texto já encontrado. A sintaxe é simples, usa-se um \ seguido pelo número do grupo, começando com 1.

**Grupos: ( )**

Grupos opcionais: ( )?

Non-capturing groups: (?:)

para definir dentro de um grupo que queremos um OU outro usamos o caractere | (pipe). Por exemplo, selecionando alfa OU beta, escrevemos (alfa|beta)

**Âncoras**

^ - é uma âncora de negação colocando no início de uma linha diz que não pode vir nada antes.

$ - semelhante ao “^” so que se usa no final para limitar que não vai vir nada depois.

\b - é uma das âncoras, chamada de word boundary. Uma âncora não seleciona um caractere, ela seleciona apenas uma posição no texto alvo.

Existe a inversão do \b, o non-word-boundary: \B (B maiúsculo)

**Classes de char: [ ]**

Apenas os caracteres \ (barra invertida), - (hífen) e ^ (circunflexo) realmente são meta-chars dentro de uma classe.

O hífen e a barra invertida já vimos na aula, por exemplo na classe: [a-z\d]

O circunflexo (^) significa negação dentro dos colchetes

**Quantifier**

? - zero ou uma vez.

\* - zero ou mais vezes.

+ - uma ou mais vezes.

{n} - exatamente n vezes.

{n,} - no mínimo n vezes.

{n,m} - no mínimo n vezes, no máximo m vezes.

**Classes de letras:**

Para descrever letras usando classes, seguem alguns exemplos:

[A-Z] - significa de A até Z, sempre maiúscula.

[a-z] - significa de a até z, sempre minúscula,

[A-Za-z] - significa A-Z ou a-z.

[abc] - significa a, b ou c.

OBS1: \s significa whitespace. A definição do \s pode variar um pouco entre as implementações, mas normalmente é um atalho para [ \t\r\n\f] onde:

O primeiro caractere é um espaço branco.

\t é um tab.

\r é carriage return.

\n é newline.

\f é form feed.