

Live de Python #19



Expressões regulares

Roteiro

- O que são
- História
- Padrões
- Caracteres
- Exemplos
- Python + regex

O que são expressões regulares?

- Método formal de especificar um padrão de texto l
- Notação para representar padrões em strings
- Composição de símbolos, caracteres e funções especiais, que, agrupados com caracteres formam uma expressão

História

- 1943: Funcionamento dos neurônios
- 1968: no qed -> ed -> grep
- 1986: Regex (Clang)

Caracteres [0]

Meta	Faz o que?
.	Qualquer caractere
[]	Lista de caracteres
?	Anterior pode existir ou não
.*	Qualquer coisa
{x}	Anterior aparece x vezes
\$	Fim da linha

Caracteres [1]

Meta	Faz o que?
+	Anterior ao menos uma vez
(xy)	Cria grupos
^	Começo da linha
\	escapa o meta
	OU
[^]	Lista negada

Listas

Meta	Faz o que?
<code>[0-9] \d</code>	acha todos os números
<code>[a-z] \w</code>	acha todas as letras minúsculas
<code>[^0-9] \D</code>	acha tudo que não seja número
<code>[^a-z] \W</code>	Acha tudo que não sejam letras

Exemplos

Padrão	Faz o que?
9?[0-9]{4}-[0-9]{4}	Número de telefone ex: 8923-9381
[M m][ilulü]l?er	Miler, muler, müller, Muler, ...
(SO so SolOslOSlos)	Buscaria variações para SO ou OS

Python + Regex = re

- Objetos
 - Um catadão de funções e constantes.
- Métodos
 - objetos com regex compiladas e que apresentam alguns métodos.

Objetos [0]

- Compile
 - Compila uma expressão e a aplica a várias strings
- Search
 - Procura uma expressão em qualquer posição de uma string
- Match[0] / Fullmatch[1]
 - Procura uma expressão no início de uma string (startswith)
 - Procura uma expressão exatamente igual
- Split
 - Dada uma expressão, fatia os dados

Objetos [1]

- Findall
 - Retorna uma lista com todos os matchs encontrados
- Finditer
 - Retorna um gerador com todos os matchs, cada um sendo um novo objeto de match.

Métodos

- search
- match
- fullmatch
- findall / finditer
- **groups**
- **pattern**

Por onde continuar? (referências)

- <https://docs.python.org/3/library/re.html>
- <http://aurelio.net/regex/guia/>
- <http://turing.com.br/material/regex/introducao.html>