Expressões Regulares

AMONTADA VALLEY

Introdução

Definições

- Basicamente, podemos dizer algo abrangente de forma específica. Definindo um padrão de busca, temos uma lista de possibilidades.

- “Como o brinquedo LEGO, várias pecinhas diferentes, cada uma com sua característica, que juntas compõem estruturas completas e podem ser arranjadas com infinitas combinações diferentes.”

- Expressões Regulares são metacaracteres que casam um padrão. É uma maneira de procurar um texto especificando padrões.

- Padrões para validar data, horário, número IP, e-mail, RG, CPF, telefones... Exemplo: **ER** para números de telefones => ([0-9]{2})? [0-9]{4}-[0-9]{4}. Casa qualquer uma das strings: “83 3224-1063”, “83 3254-3421”, “82 2343-1212”...

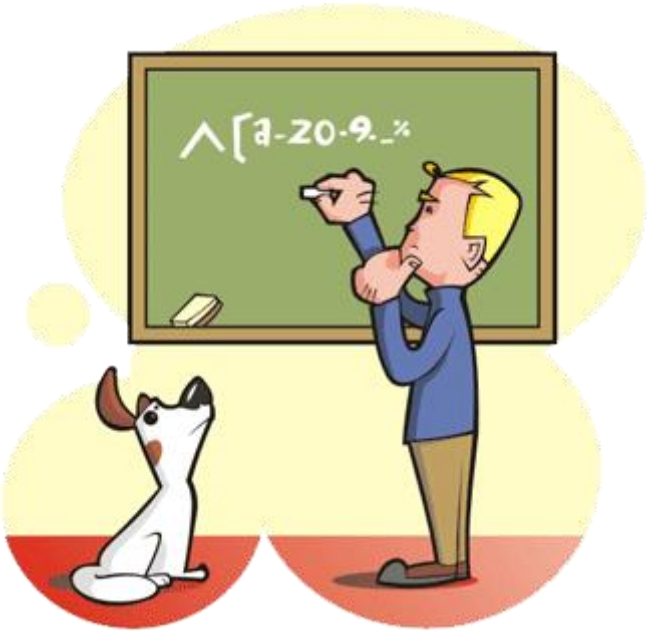
História

• O termo deriva do que se chamou de álbegra de conjuntos regulares (*"regular sets"*), do matemático Stephen Cole Kleene.

• 1968, Ken Thompson implementava no *qed* um comando de contexto que aceitava expressões regulares. Sua sintaxe? g/RE/p (Global Regular Expression Print)... Mais tarde deu origem ao famoso grep nos sistemas Unix.

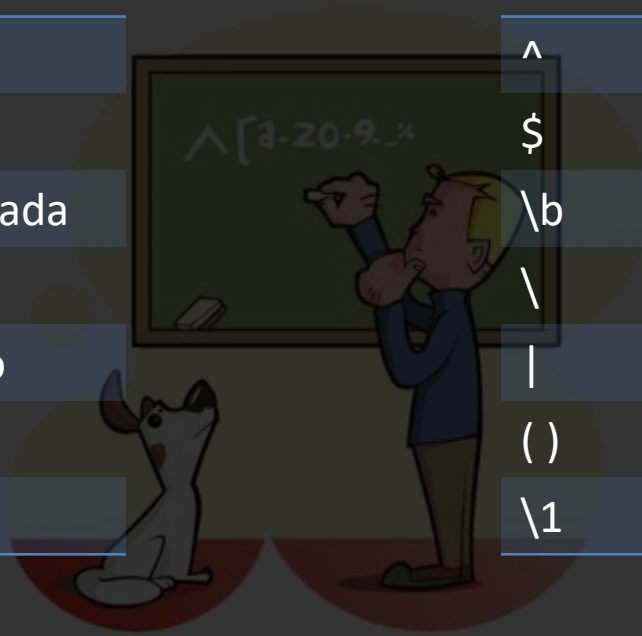
• 1986, criado um pacote pioneiro em C, chamado regex e de graça! Daí a norma IEEE POSIX 1003.2 (POSIX.2) padroniza expressões regulares.

Os Metacaracteres



Sopa de letrinhas

Metacaractere	Nome
.	Ponto
[ ]	Lista
[^]	Lista negada
?	Opcional
*	Asterisco
+	Mais
{ }	Chaves



Metacaractere	Nome
^	Circunflexo
$	Cifrão
\b	Borda
\	Escape
|	Ou (pipe)
( )	Grupo
\1	Retrovisor

Sopa de letrinhas – exemplos

Miller, Miler, Müler, Müller, Mueler, Mueller:

M(ü|i|ue)ll?er

Siglas de 4 letras, começando com UF:

UF[A-Z]{2}

Tags HTML , , , , <u>, </u>:

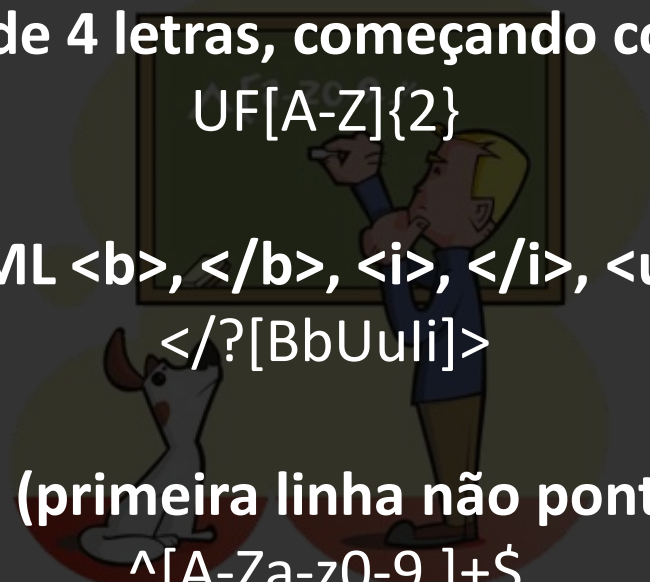
</?[BbUuIi]>

Títulos (primeira linha não pontuada):

^[A-Za-z0-9]+$

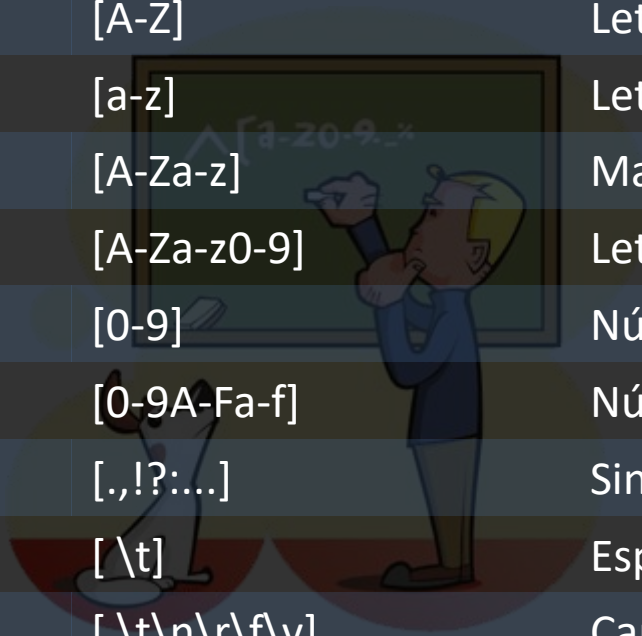
Palavras:

\b[A-Za-z]+\b



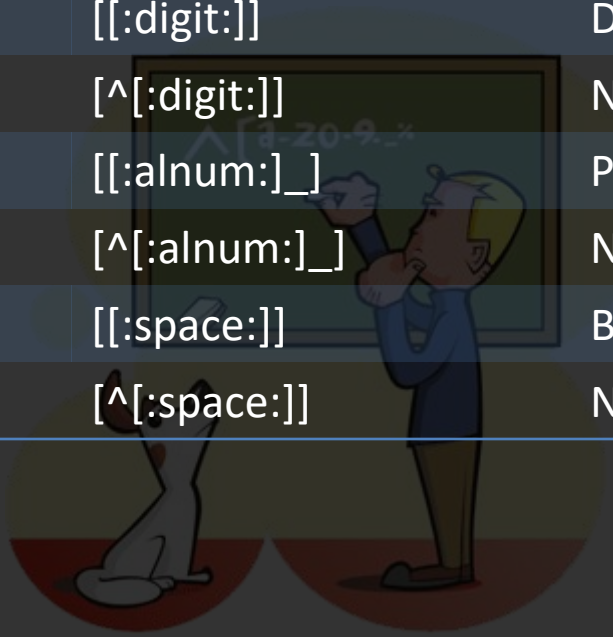
Padrões POSIX

Classe POSIX	Similar	Significa
[:upper:]	[A-Z]	Letras maiúsculas
[:lower:]	[a-z]	Letras minúsculas
[:alpha:]	[A-Za-z]	Maiúsculas/Minúsculas
[:alnum:]	[A-Za-z0-9]	Letras e números
[:digit:]	[0-9]	Números
[:xdigit:]	[0-9A-Fa-f]	Números Hexadecimais
[:punct:]	[.,!?:...]	Sinais de pontuação
[:blank:]	[\t]	Espaço e TAB
[:space:]	[\t\n\r\f\v]	Caracteres brancos

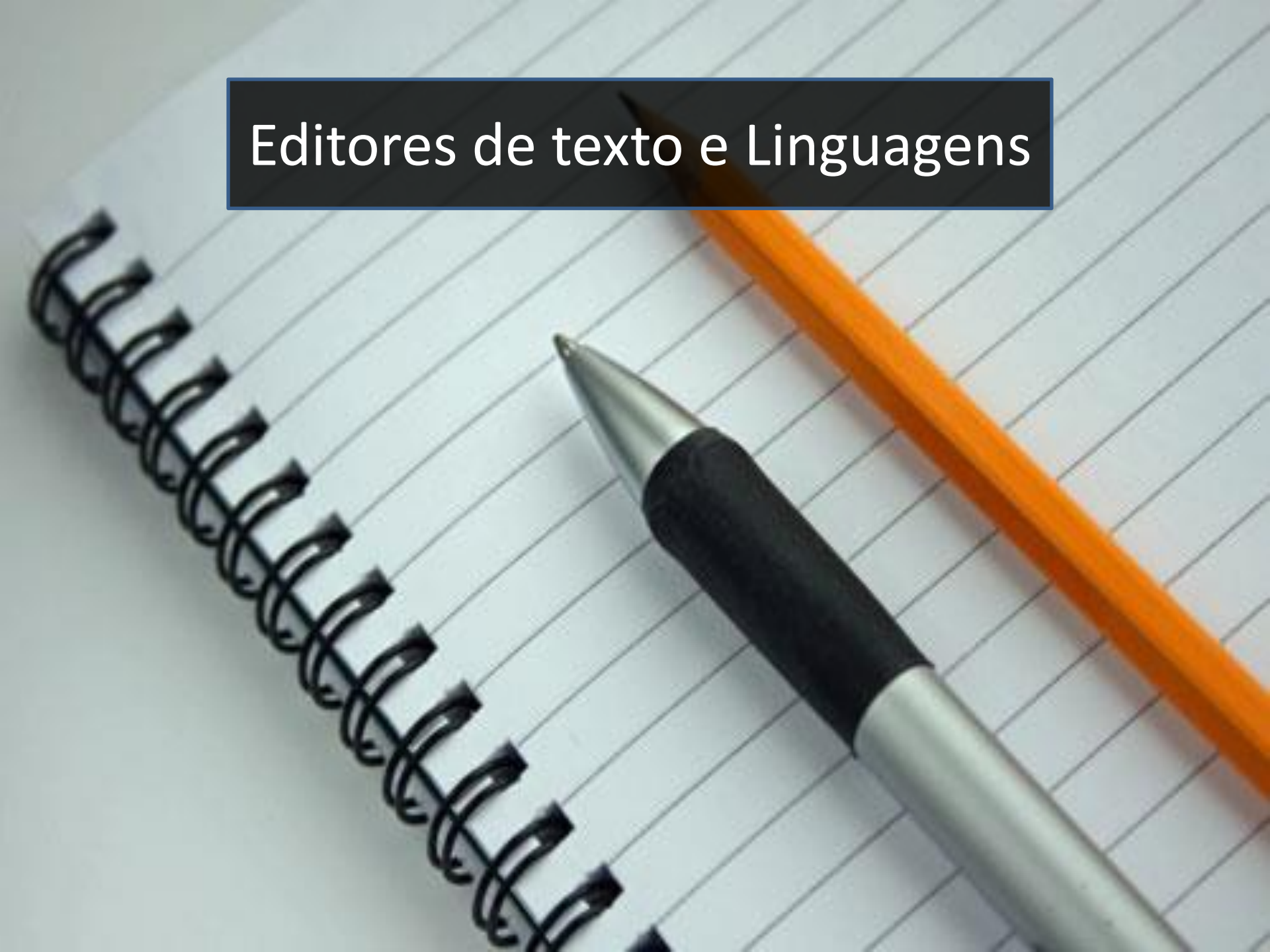


(Alguns m|M)etacaracteres tipo barra-letra

Classe POSIX	Similar	Significa
\d	[[:digit:]]	Dígito
\D	[^[:digit:]]	Não-dígito
\w	[[:alnum:]_]	Palavra
\W	[^[:alnum:]_]	Não-palavra
\s	[[:space:]]	Branco
\S	[^[:space:]]	Não-branco



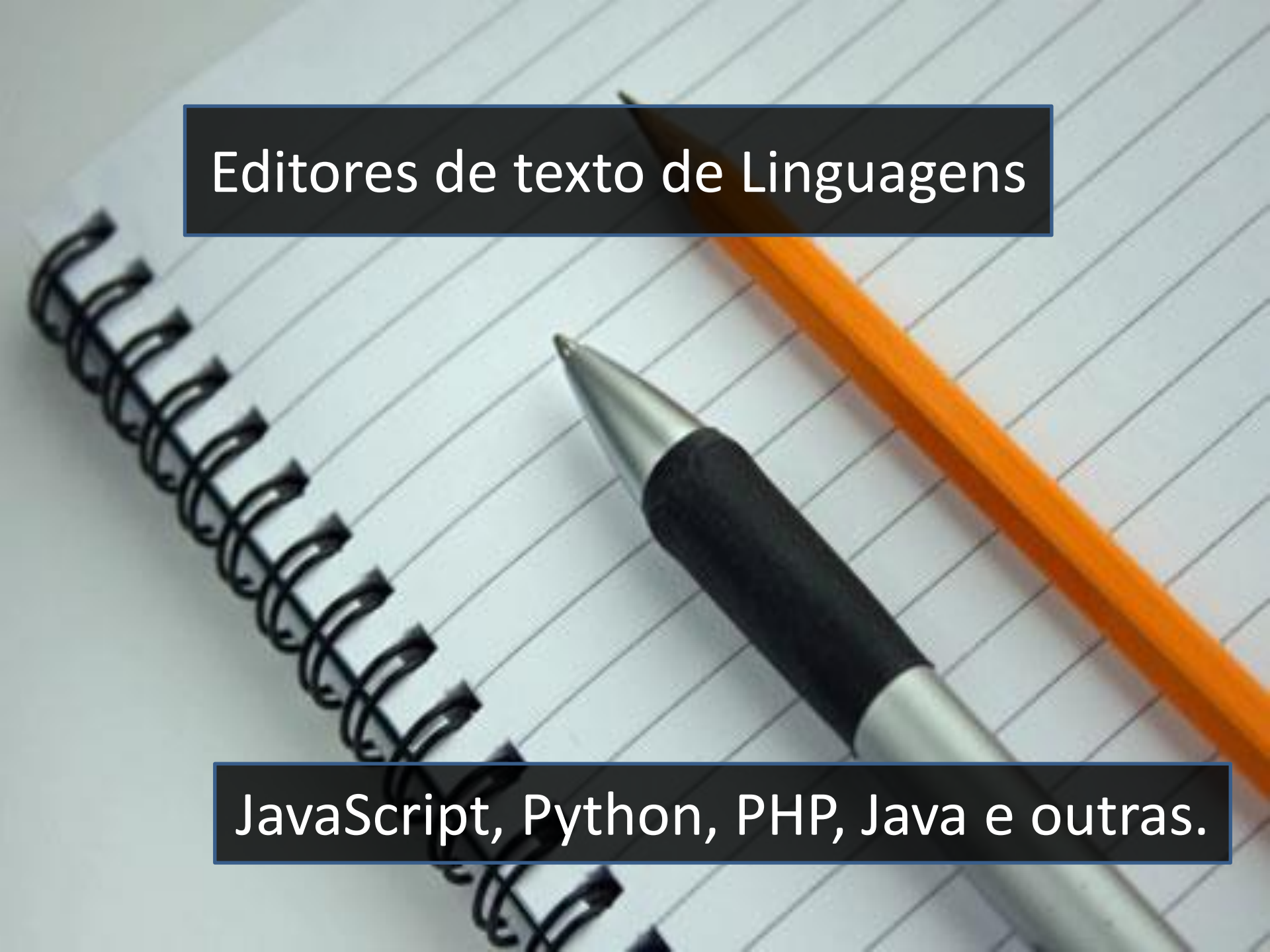
Editores de texto e Linguagens



Observações

- Não há classes do POSIX e não suporta acentuação.
- Solução? [A-Za-zÀ-u]

•Quebras de linha e paragráfos podem ser identificados com \n, assim as ERs podem casar múltiplas linhas de uma vez.


Editores de texto de Linguagens
JavaScript, Python, PHP, Java e outras.

Linguagens de Programação

- Na J2SE aparecem no pacote java.util.regex, além de algumas melhorias na classe java.lang.String.

- No Python deve ser criado um “import re” para utilizar ReGex nos códigos.

- No PHP e JavaScritp são métodos nativos, não necessitam de nada especial para utilização.

- No Javascript são baseadas em dois elementos: um padrão e outro o modificador. A sintax é bastante simples /padrão/modificador. /Valley/i por exemplo pode ser utilizada para encontrar a palavra “valley” e a letra “i” indica que não precisa se preocupar com maiúsculas e minúscula.

Exemplos

Data (dd/mm/aaaa)

```
[0-9]{2}/[0-9]{2}/[0-9]{4}
[0123][0-9]/[01][0-9]/[12][0-9]{3}
([012][0-9]|3[0-1])/([0[0-9]|1[012])/[12][0-9]{3}
(0[1-9]|[12][0-9]|3[01])/0[0-9]|1[012])/[12][0-9]{3}
```

Horário (hh:mm)

```
[0-9]{2}:[0-9]{2}
[012][0-9]:[0-5][0-9]
([01][0-9]|2[0-3]):[0-5][0-9]
```

Email (email@company.com)

```
[^@]+@[^@]+.[.*]
[A-Za-z0-9_.-]+@[A-Za-z0-9_]+\.[a-z]+
[A-Za-z0-9_.-]+@[A-Za-z0-9_]+\.[a-z]{2,3}
```

Bibliografia

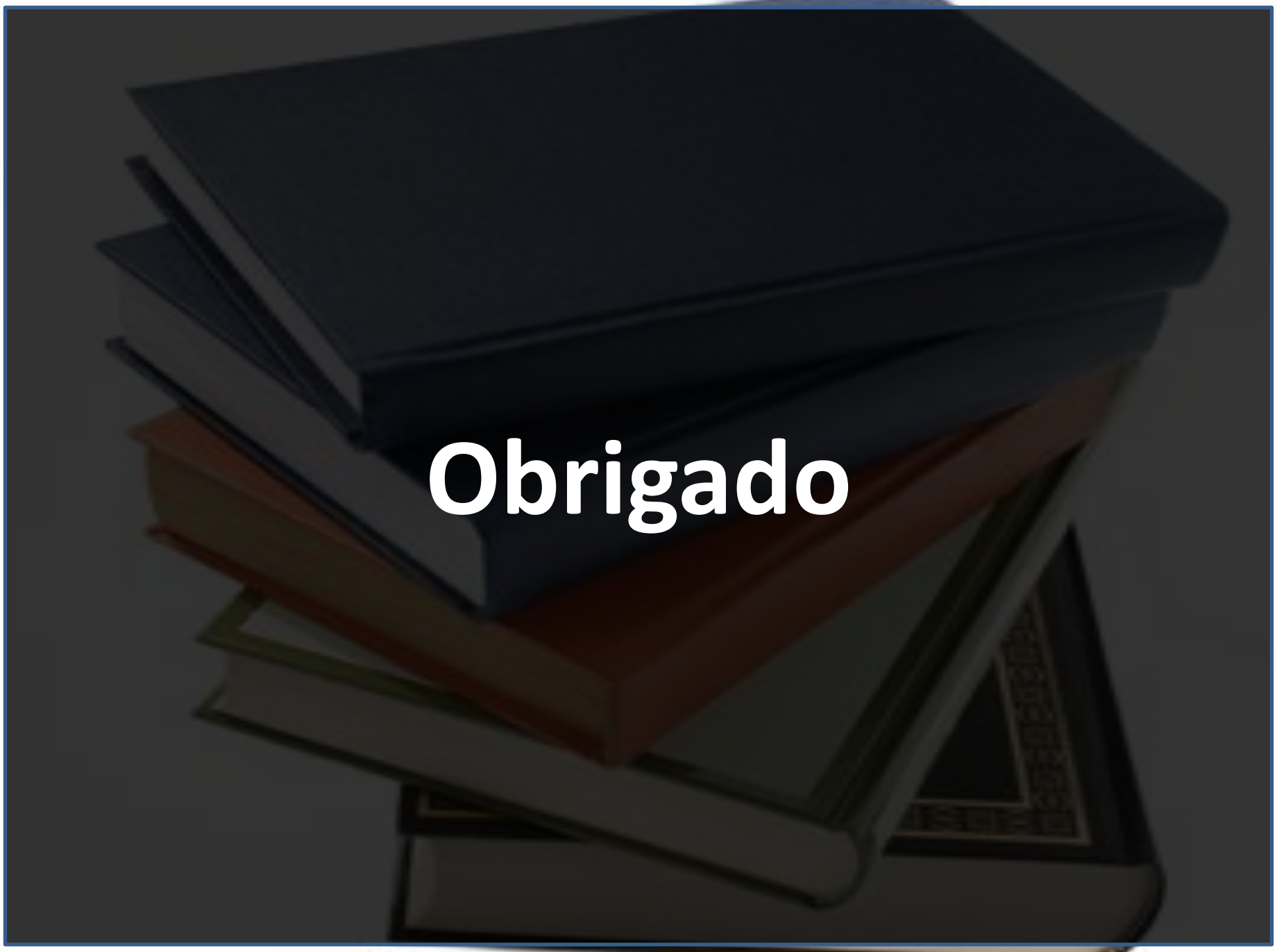
Jargas, Aurélio Marinho. **Expressões Regulares**: Uma abordagem divertida. Ed. Novatec, 3ª edição. São Paulo, 2009.

Referências

Blog do Autor do livro Expressões Regulares:
http://aurelio.wordpress.com/2009/07/17/10anos-expressoes-regulares/
Site oficial do livro Expressões Regulares: http://www.piazinho.com.br/
Expressão regular: http://pt.wikipedia.org/wiki/Express%C3%A3o_regular

+info

Expressões regulares com PHP e Python:
http://blog.rafaelcapucho.com/expressoes-regulares/expressoes-regulares-com-php-e-python-na-pratica.html
Expressões regulares com Ruby:
http://wiki.cercomp.ufg.br/Equipe_Web/RoR/Express%C3%A3o_Regular
Mini-Curso de Expressões Regulares:
http://www.brasiltech.net/informatiquez/2008/04/09/mimi-curso-de-expressoes-regulares-entenda-como-funciona-parte-1/
Exemplos do livro Expressões Regulares:
http://www.piazinho.com.br/exemplos.html


Obrigado