Expressões Regulares na Prática!

Gustavo Yudi Bientinezi Matsuzake

Universidade Tecnológica Federal do Paraná
Coders UTFPR

matsuzake@alunos.utfpr.edu.br
`https://github.com/yudi-matsuzake/regex-coders`

10 de Novembro de 2015

Sumário

Definição

Expressão Regular é:

- Uma "linguagem de programação"[1];
- Intimamente relacionada aos DFA, que são *autômatos*, analisadores sintáticos[1];
- Uma maneira sucinta e finita de representar uma linguagem infinita - regular;
- É um método formal de se especificar um padrão de texto[2].

Mesmo, Eu.

"Uma imagem vale mais do que mil palavras. Uma Expressão Regular[1] vale infinitas."

[1] bombeável

Observação importante:

Diferença da teoria e da prática.

Vamos deixar claro nosso escopo!

Para a felicidade de alguns, infelicidade de outros...

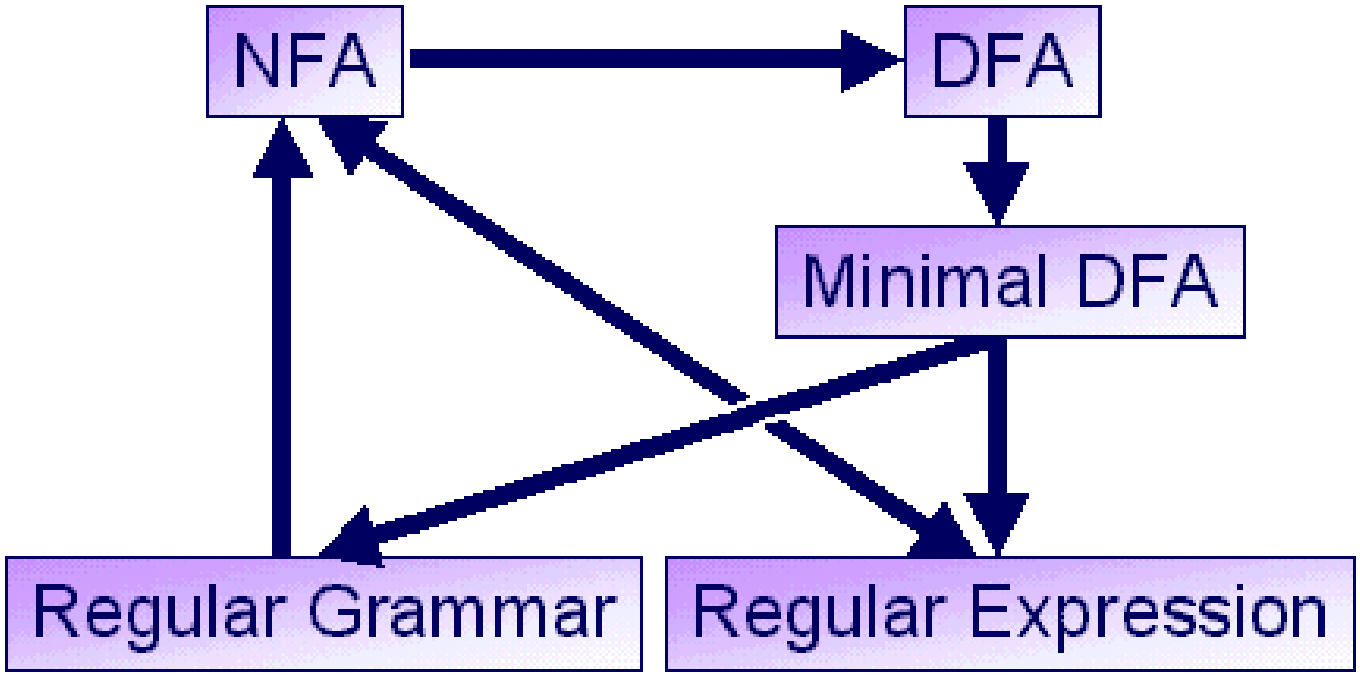
- Nosso escopo nessa apresentação é mais prático e não teórico;
- Vamos aprender como usar;
- Vamos aprender onde usar;
- Vamos aprender porque usar.

Nada de lógica

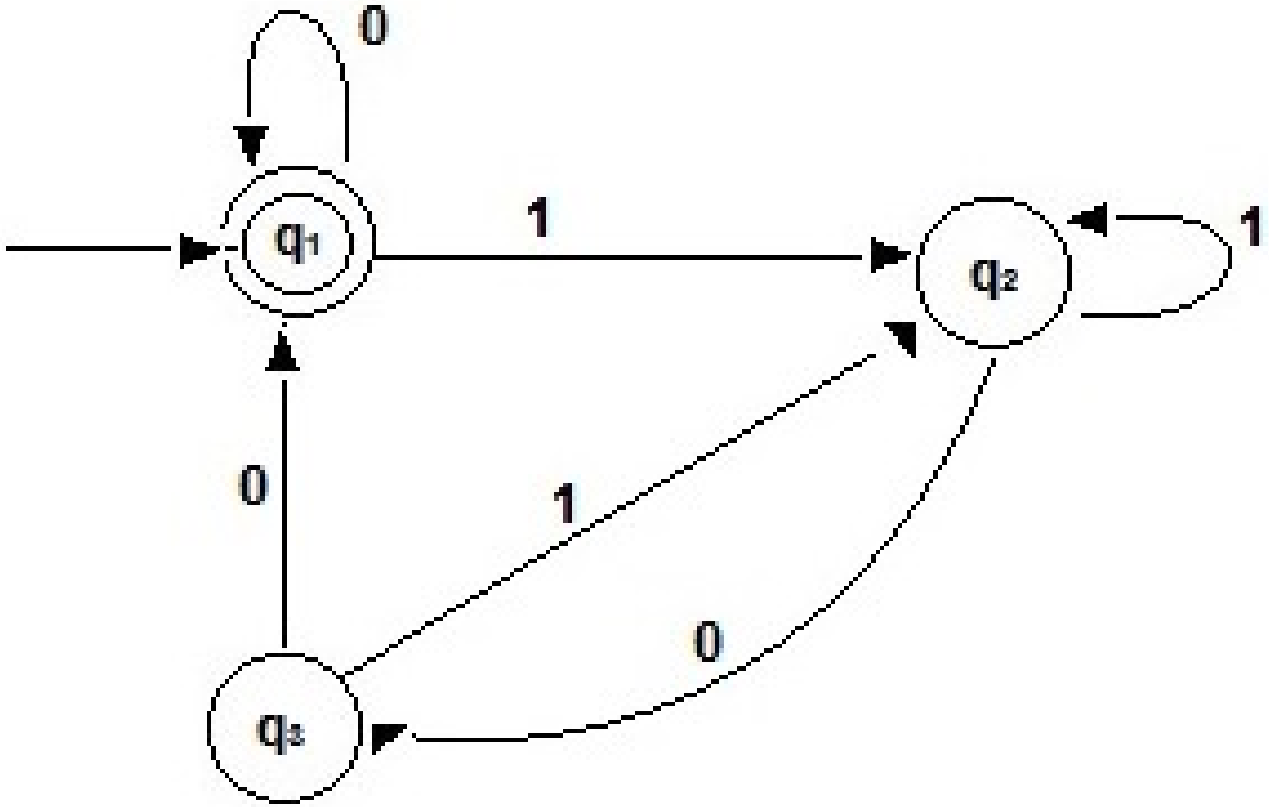
List of Identities:

1. $P \Leftrightarrow (P \vee P)$ ----- idempotence of $\vee$
2. $P \Leftrightarrow (P \wedge P)$ ----- idempotence of $\wedge$
3. $(P \vee Q) \Leftrightarrow (Q \vee P)$ ----- commutativity of $\vee$
4. $(P \wedge Q) \Leftrightarrow (Q \wedge P)$ ----- commutativity of $\wedge$
5. $[(P \vee Q) \vee R] \Leftrightarrow [P \vee (Q \vee R)]$ ----- associativity of $\vee$
6. $[(P \wedge Q) \wedge R] \Leftrightarrow [P \wedge (Q \wedge R)]$ ----- associativity of $\wedge$
7. $\neg (P \vee Q) \Leftrightarrow (\neg P \wedge \neg Q)$ ----- DeMorgan's Law
8. $\neg (P \wedge Q) \Leftrightarrow (\neg P \vee \neg Q)$ ----- DeMorgan's Law
9. $[P \wedge (Q \vee R] \Leftrightarrow [(P \wedge Q) \vee (P \wedge R)]$ ----- distributivity of $\wedge$ over $\vee$
10. $[P \vee (Q \wedge R] \Leftrightarrow [(P \vee Q) \wedge (P \vee R)]$ ----- distributivity of $\vee$ over $\wedge$
11. $(P \vee True) \Leftrightarrow True$
12. $(P \wedge False) \Leftrightarrow False$
13. $(P \vee False) \Leftrightarrow P$
14. $(P \wedge True) \Leftrightarrow P$
15. $(P \vee \neg P) \Leftrightarrow True$
16. $(P \wedge \neg P) \Leftrightarrow False$
17. $P \Leftrightarrow \neg (\neg P)$ ----- double negation
18. $(P \rightarrow Q) \Leftrightarrow (\neg P \vee Q)$ ----- implication
19. $(P \leftrightarrow Q) \Leftrightarrow [(P \rightarrow Q) \wedge (Q \rightarrow P)]$----- equivalence

Nada de transição de DFA para NFA



Não vamos aprender autômatos



Isso vai fazer parte da sua realidade

```
^*[A-Za-z0-9_]+:(.*)$
```

Isso não vai te dar mais pânico!

```
([0-9]{1,3}\.){2}[0-9]{1,3}-[0-9]{2}
```

Isso será fácil de entender (:

```
[-+]?[0-9]{1,3}(\.[0-9]{3})?(,[0-9]{2})?
```

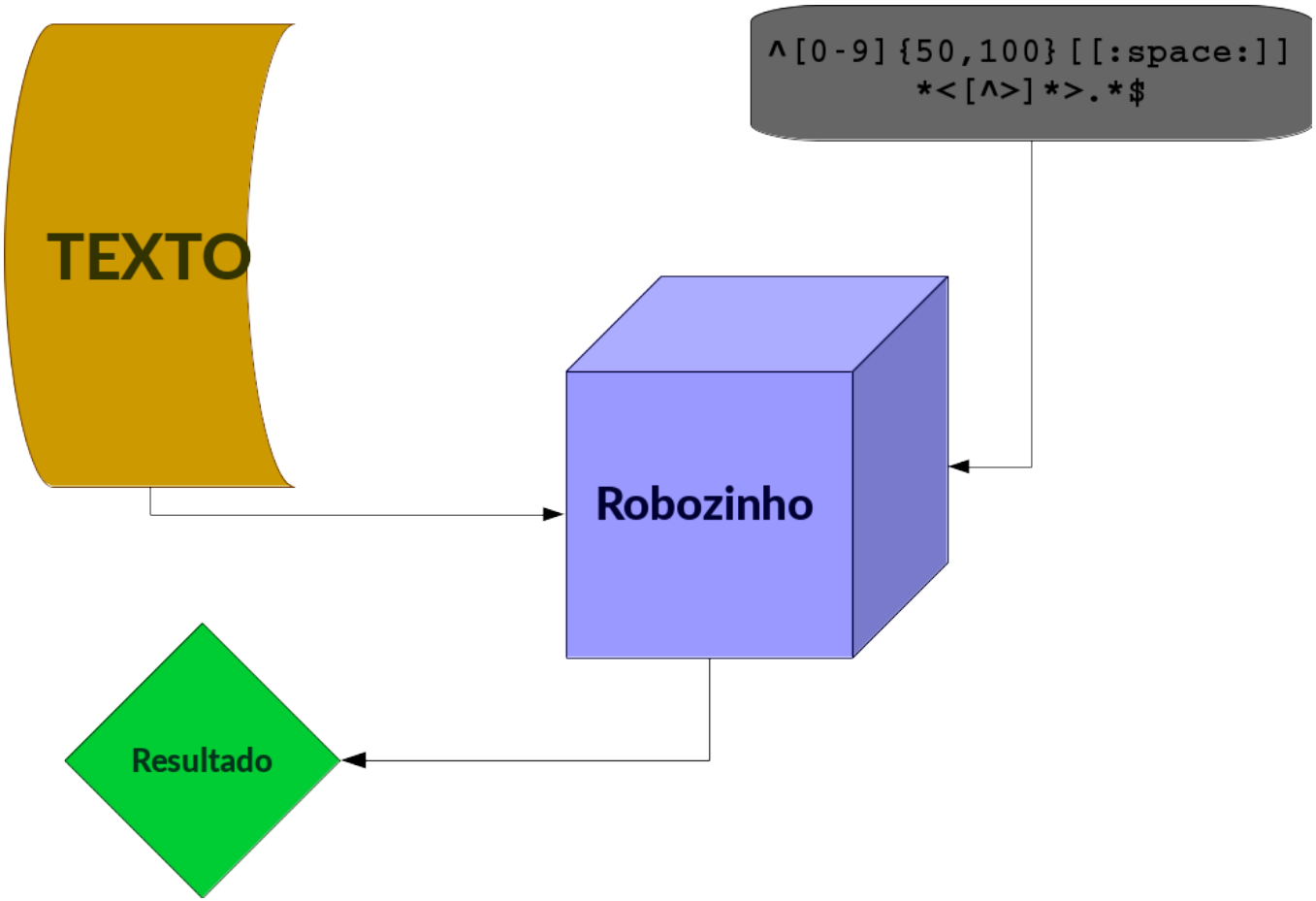


Figura: Como vamos abordar as RE

[Expressões Regulares] Uma abordagem divertida

Figura: Essa apresentação foi inspirada, causada e é sobre esse livro.

Você sabe o que são os metacaracteres?

- Se você não conhecia essas belezinhas de *expressões regulares* você só utilizou caracteres literais durante toda sua vida;
- Metacaracteres, você, você metacaracteres;
- Como isso vai fazer diferença na minha vida?
- Diferença dos robozinhos.

Figura: Expressões Regulares salvando o dia

Tipos

Representante e Quantificador

Ponto

.

O ponto

O ponto casa com TUDO.

i.e., o ponto casa com qualquer caractere!

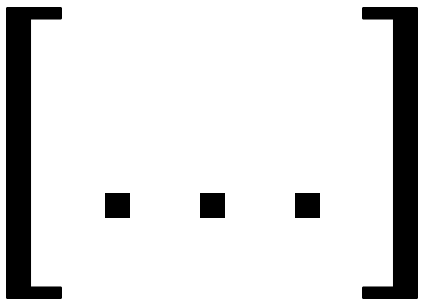
Pergunta: o ponto casa com um ponto?

Figura: O ponto é o coringa solitário.

Casamento do ponto

Expressão	Casa com...
n.o	nao, não, n40, noo...
.u	au, bu, du...
12.45	12:45, 12 45, 12.45...

Lista



A lista

- A lista casa com TODOS os caracteres dentro dela.
- **Exemplos**: [abc], [123], [letras]...
- **Dentro da lista todo mundo é LITERAL.**
  - O ponto dentro da lista é literal;

Mentiroso!

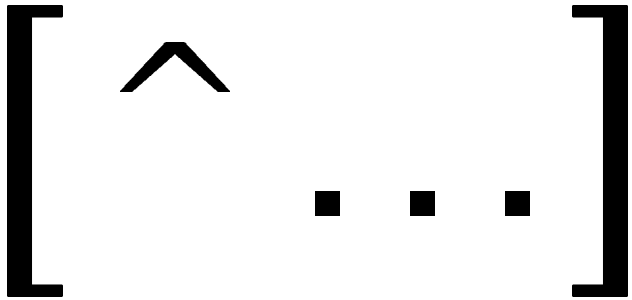
- Lembra que eu falei que todo mundo dentro lista era literal?

  Então... **EU MENTI MUAHAHAHA**.

Mentiroso (nem tanto, vai)!

- Lembra que eu falei que todo mundo dentro lista era literal? Então...
  - Todo caractere que se acha fora da lista, lá dentro não serve pra nada ou tem outro significado;
  - É como se ela tivesse um mundinho dentro dela, só dela, onde ela dita as regras.

Lista Negada



A lista negada

- A lista negada casa com TODOS os caracteres que **NÃO** estão dentro dela.
- Ou seja, ela não é tão seletiva quanto a lista, tampouco liberal quanto o ponto. Ela sabe com quem não quer casar;
- **Exemplos**: [^abc], [^123], [^letras]...

Intervalos

- Como você faria uma expressão regular para casar com dois números consecutivos?

Intervalos

- Como você faria uma expressão regular para casar com dois números consecutivos?
  - "**[0123456789][0123456789]**"?

Intervalos

- O traço (-) é um operador não literal dentro da lista;
- Ele foi criado para representar intervalos:
  - [0-9] é igual a [0123456789];
  - [a-z] é igual a [abcdefghijklmnopqrstuvwxyz];
  - [A-Z] é igual a [ABCDEFGHIJKLMNOPQRSTUVWXYZ];
  - [0-9a-zA-Z]
  - [-~]

Intervalos

- O traço (-) é o único operador não literal dentro da lista;
- Ele foi criado para representar intervalos:
  - [0-9] é igual a [0123456789];
  - [a-z] é igual a [abcdefghijklmnopqrstuvwxyz];
  - [A-Z] é igual a [ABCDEFGHIJKLMNOPQRSTUVWXYZ];
  - [0-9a-zA-Z];
  - [-~] ???

Exceções

- E se eu quiser colocar um traço (-) na lista?
- E se eu quiser colocar os colchetes ([e]) na lista?
- E se eu quiser colocar um circunflexo na lista?
- E se...

Exceções

- **E se eu quiser colocar um traço (-) na lista?**
  - Devemos color o traço no começo ou no final da lista;
  - [-0-9] → Casa com 0 a 9 e traço;
  - [a-f-] → Casa com a a f e traço;
- **E se eu quiser colocar os colchetes ([e]) na lista?**
- **E se eu quiser colocar um circunflexo na lista?**
- **E se...**

Exceções

- **E se eu quiser colocar um traço (-) na lista?**
- **E se eu quiser colocar os colchetes ([e]) na lista?**
  - O "colchete abrir"podemos colocar em qualquer lugar:
    - [[0-9];
    - [0-4[*.ç];
  - O "colchete de fechar"deve ser o primeiro da lista;
    - []0-9] → Casa com 0 a 9 ou com].
    - [0-9]] → Tá errado.
- **E se eu quiser colocar um circunflexo na lista?**
- **E se...**

Exceções

- **E se eu quiser colocar um traço (-) na lista?**
- **E se eu quiser colocar os colchetes ([e]) na lista?**
- **E se eu quiser colocar um circunflexo na lista?**
  - [a-z^]→casa com caracteres de a a z e ^;
  - [^a-z]→lista negada: casa com qualquer coisa que não seja de a a z.
- **E se...**

Exceções

- **E se eu quiser colocar um traço (-) na lista?**
- **E se eu quiser colocar os colchetes ([e]) na lista?**
- **E se eu quiser colocar um circunflexo na lista?**
- **E se...**
  - Mais alguma coisa/dúvida?

Casamento das listas

Expressão	Casa com...
n[aãAÃ]o	nao, não, nAo, nÃo
12[:.]45	12:45, 12.45, 12 45
funcao.[ch]	funcao.c, funcao.h
[][-]	???

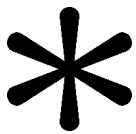
Casamento das listas

Expressão	Casa com...
n[aãAÃ]o	nao, não, nAo, nÃo
12[:.]45	12:45, 12.45, 12 45
funcao.[ch]	funcao.c, funcao.h
[][-]	[,], -

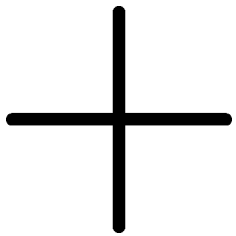
Opcional

?

Asterisco



Mais



Chaves

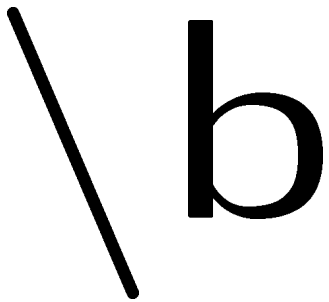
$$\{m,n\}$$

Circunflexo

Cifrão

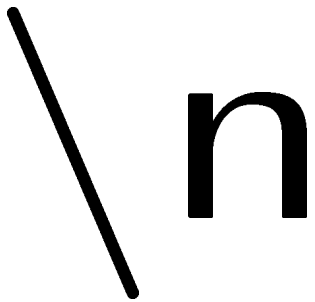
$

Borda

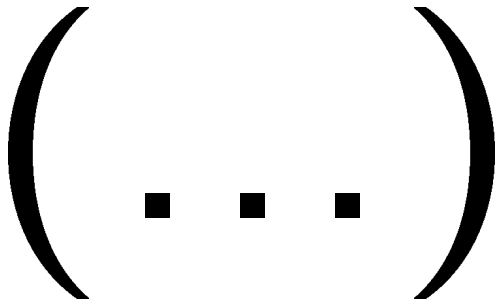


Ou

Escape



Grupo



Retrovisor

\1...\9

Referências

John Hopcroft, Jeffrey Ullman, and Rajeev Motwani.
Introduction to Automata Theory Languages and Computation.
Pearson Education, 2 edition, 2001.

Aurelio Marinho Jargas.
Expressões Regulares - Uma abordagem Divertida.
Novatec, www.aurelio.net, 4 edition, 2012.
Obrigado Aurelio por apresentar esse mundo bem louco e empolgante do software livre, por me fazer se apaixonar por shell script, vim e regex. SÉDimais (você é demais!).

Acabou :(

Dúvidas?

Expressões Regulares na Prática!

Gustavo Yudi Bientinezi Matsuzake

Universidade Tecnológica Federal do Paraná
Coders UTFPR

matsuzake@alunos.utfpr.edu.br
`https://github.com/yudi-matsuzake/regex-coders`

10 de Novembro de 2015