



Universidade Federal do Pará
Instituto de Ciências Exatas e Naturais
Programa de Pós-Graduação em Ciência da Computação
Disciplina: Teoria da Computação
Professor: Reginaldo Santos

Trabalho sobre Expressão Regular

1. [4.0 pontos] O preenchimento de um conjunto de campos em um formulário é uma tarefa comum que se faz presente em diversas aplicações para *web*, *mobile* e *desktop*. Cada campo, normalmente, necessita obedecer um conjunto de regras para que sua validação retorne sucesso. Por exemplo, em um campo de e-mail, tipicamente brasileiro, espera-se obter cadeias na forma

$$(a|b|...|z)^+@(a|b|...|z)^+(\.com|.com.br)$$

As expressões regulares são técnicas robustas e rápidas para verificar se uma determinada cadeia, i.e., a *string* digitada pelo usuário, pertence a uma determinada linguagem, i.e., obedece ao conjunto de regras sobre o alfabeto do campo. Neste contexto, desenvolva, na linguagem de programação de sua preferência, um conjunto de **máscaras de validação**, por meio de expressões regulares, que obedecem as especificações prévias de cada campo abaixo. Considere os seguintes alfabetos $\Sigma = \{a, b, c, \dots, z\}$, $\Gamma = \{A, B, C, \dots, Z\}$ e $N = \{0, 1, 2, \dots, 9\}$.

- **Nome, nome do meio e sobrenome:**

1. Nome, nome do meio e sobrenome devem vir separados por um espaço apenas
2. O nome do meio é opcional
3. Nome e sobrenome devem ser ambos não vazios
4. Não deve aceitar caracteres especiais ou números
5. O primeiro símbolo do nome e sobrenome, e do nome do meio se existir, deve ser do alfabeto Γ e os outros símbolos devem ser do alfabeto Σ

Ex. de sentenças aceitas: Ada Lovelace, Alan Turing, Stephen Cole Kleene

Ex. de cadeias rejeitadas: 1Alan, Alan, A1an, A1an Turing, Alan turing

- **E-mail:**

1. Sentenças devem conter um, e apenas um, símbolo “@”
2. Excetuando o símbolo “@”, as sentenças possuem apenas símbolos de Σ



Universidade Federal do Pará
Instituto de Ciências Exatas e Naturais
Programa de Pós-Graduação em Ciência da Computação
Disciplina: Teoria da Computação
Professor: Reginaldo Santos

3. Sentenças não devem começar com o símbolo “@”
4. Sentenças devem terminar com a subcadeia “.com.br” ou “.br”
5. Sentenças devem ter, pelo menos, um símbolo de Σ entre o símbolo “@” e a subcadeia “.com.br” ou a subcadeia “.br”

Ex. de sentenças aceitas: a@a.br, divulga@ufpa.br, a@a.com.br

Ex. de cadeias rejeitadas: @, a@.br, @a.br, T@teste.br, a@A.com.br

- **Senha:**

1. Sentenças podem conter símbolos de $\Sigma \cup \Gamma \cup N$
2. Sentenças devem ter pelo menos um símbolo de Γ e outro de N
3. Sentenças devem ter comprimento igual a 8

Ex. de sentenças aceitas: 518R2r5e, F123456A, 1234567T, ropsSoq0

Ex. de cadeias rejeitadas: F1234567A, abcdefgH, 1234567HI

Obs.: nesta linguagem é permitido utilizar recursos da linguagem de programação (e.g., size, length ou *lookahead*) para descobrir se a cadeia tem comprimento 8.

- **CPF:**

1. Sentenças devem ter o formato **xxx.xxx.xxx-xx**, onde $x \in N$

Ex. de sentenças aceitas: 123.456.789-09, 000.000.000-00

Ex. de cadeias rejeitadas: 123.456.789-0, 111.111.11-11

- **Telefone:**

1. Sentenças devem ter um dos formatos seguintes

(xx) 9xxxx-xxxx

(xx) 9xxxxxxxx

xx 9xxxxxxxx

onde $x \in N$.

Ex. de sentenças aceitas: (91) 99999-9999, (91) 999999999, 91 999999999

Ex. de cadeias rejeitadas: (91) 59999-9999, 99 99999-9999, (94)95555-5555



Universidade Federal do Pará
Instituto de Ciências Exatas e Naturais
Programa de Pós-Graduação em Ciência da Computação
Disciplina: Teoria da Computação
Professor: Reginaldo Santos

- **Data e horário:**

1. Sentenças devem ter o formato **dd/mm/aaaa hh:mm:ss**, onde $d, m, a, h, m, s \in \mathbb{N}$.
Ex. de sentenças aceitas: 31/08/2019 20:14:55, 99/99/9999 23:59:59
Ex. de cadeias rejeitadas: 99/99/9999 3:9:9, 9/9/99 99:99:99, 99/99/999903:09:09

- **Número real com ou sem sinal:**

1. Sentenças devem começar com um dos símbolos do alfabeto $\{+, -, \epsilon\}$
 2. Em seguida, as sentenças devem ter, pelo menos, um símbolo do alfabeto \mathbb{N}
 3. Em seguida, as sentenças devem ter, exatamente, um símbolo separador “.”
 4. Em seguida, as sentenças devem finalizar com, pelo menos, um símbolo de \mathbb{N}
 5. Exceção: números sem a parte fracionária também devem ser aceitos
- Ex. de sentenças aceitas: -25.467, 1, -1, +1, 64.2
Ex. de cadeias rejeitadas: 1., .2, +64,2



Universidade Federal do Pará
Instituto de Ciências Exatas e Naturais
Programa de Pós-Graduação em Ciência da Computação
Disciplina: Teoria da Computação
Professor: Reginaldo Santos

2. [6.0 pontos] Implemente, na linguagem de programação de sua preferência, os arranjos familiares solicitados nos itens abaixo sobre o alfabeto $\Sigma = \{H, M, h, m\}$, com significado:

- **H** representa um homem;
- **M** representa uma mulher;
- **h** representa um filho do sexo masculino (natural ou adotado);
- **m** representa uma filha do sexo feminino (natural ou adotado);
- A posição relativa de uma letra em relação às demais indica a idade relativa daquele membro da família em relação aos demais (os mais novos estão sempre à direita).

Exemplo: a cadeia “MHhmm” representa uma família com um casal heterossexual em que a mulher é mais velha que o homem. Além disso, esse casal possui três filhos, um homem e duas mulheres, sendo que o filho homem é o mais velho dos três.

- a) Casais heterossexuais mais velhos que os filhos com pelo menos duas filhas mulheres, ou pelo menos um filho homem, ou ainda pelo menos dois filhos homens e uma filha mulher.
- b) Casais heterossexuais mais velhos que os filhos e com uma quantidade ímpar de filhas mulheres.
- c) Casais heterossexuais mais velhos que os filhos, com a filha mais velha mulher e o filho mais novo homem.
- d) Casais homossexuais mais velhos que os filhos, com pelo menos seis filhos, em que os dois primeiros filhos formam um casal e os últimos também.
- e) Casais homossexuais mais velhos que os filhos, em que o sexo dos filhos é alternado conforme a ordem de nascimento.
- f) Casais homossexuais mais velhos que os filhos, com qualquer quantidade de filhos homens e mulheres, mas que não tiveram dois filhos homens consecutivos.
- g) Arranjo de no mínimo $x \in \mathbb{N}$ e no máximo $y \in \mathbb{N}$, com $x > 0, y > 0$, e $x \leq y$, de adultos (Hs ou Ms) mais velhos que os filhos, com qualquer quantidade de filhos homens e mulheres, mas que os três filhos mais novos não foram homens.



Universidade Federal do Pará
Instituto de Ciências Exatas e Naturais
Programa de Pós-Graduação em Ciência da Computação
Disciplina: Teoria da Computação
Professor: Reginaldo Santos

Instruções para a entrega do trabalho

Deve ser entregue um relatório técnico, em formato de artigo SBC¹, juntamente com código-fonte de cada questão. O relatório deve conter, mas não limitado a:

- *Abstract* e Resumo
- Introdução
- Materiais e métodos
- Testes experimentais
- Comentários finais

Anexar o código-fonte, como Apêndice, ao final do documento². O relatório deve ter no mínimo 10 páginas (não contar as páginas contendo o código-fonte). Use sua criatividade para organizar o relatório em subseções e expressar o que foi feito no decorrer do trabalho, incluindo uma justificativa para a linguagem de programação escolhida.

O relatório em formato de artigo deve ser convertido para a extensão PDF e ser submetido via SIGAA antes da data limite. **Não serão aceitos trabalhos após a data limite pré-definida.**

¹ Template: <https://www.sbc.org.br/documentos-da-sbc/send/169-templates-para-artigos-e-capitulos-de-livros/878-modelosparapublicaodeartigos>

² Alternativamente, o código-fonte pode ser disponibilizado na nuvem e o link deve ser adicionado no relatório técnico.