



## Laboratório

Curso Superior em Engenharia de Computação  
Disciplina de Linguagem Formais e Autômatos  
Professor Gomide

Chegou a hora de criarmos aplicações reais utilizando o poder das expressões regulares.  
Nessa proposta faremos dois estudos de caso.

O código produzido nos dois estudos, serão aceitos apenas se forem escritos em: Java;  
JavaScript; ou Python.

## Estudo de Caso A

Crie um programa que acesse o código-fonte do site de blogs e colunas da UOL:

<https://noticias.uol.com.br/blogs-e-colunas/>

A partir do código-fonte extraído deste domínio crie uma expressão regular que capture as principais notícias presentes nessa página. Para cada notícia extraia o título da notícia, o texto e o autor.

The screenshot shows the UOL website interface. At the top, there's a navigation bar with the UOL logo and links for Jogos, Eleições, Canal UOL, and Colunas. Below this, the main heading 'COLUNAS' is displayed. A large blue letter 'A' is followed by a paragraph: 'qui você encontra todas as colunas do UOL, com análises e reportagens de mais de 130 personalidades sobre economia, esportes, política, saúde, diversidade, direitos da mulher e mais. Confira!'. Below this, there's a section titled 'Todas as notícias' which contains a grid of news items. Each item has a small image, a title, and a brief description. A legend on the left side of the grid identifies the fields: 'Título' (Title), 'Texto' (Text), and 'Autor' (Author). The first item in the grid is titled 'HISTÓRIAS DO MAR' and describes a historical ship sinking. The second item is titled 'POLE POSITION' and mentions Colapinto. The third item is titled 'MAURO CEZAR PEREIRA' and mentions Chile and Peru. The fourth item is titled 'VOZ DOS OCEANOS' and mentions plastic bottles.

Título	Texto	Autor
HISTÓRIAS DO MAR	Navio histórico vai afundar e se tornará maior recife artificial do mundo	Jorge de Souza
POLE POSITION	Colapinto diz que é mais difícil ser sul-americano na F1 e elogia Bortoleto	Julianne Cerasoli
MAURO CEZAR PEREIRA	Se não vencer Chile e Peru, o Brasil deve pensar com carinho na repescagem	Mauro Cezar Pereira
VOZ DOS OCEANOS	Mais de 50% das garrafas plásticas nas nossas praias vêm de outros países	Vilfredo Schurmann



## Laboratório

Curso Superior em Engenharia de Computação  
Disciplina de Linguagem Formais e Autômatos  
Professor Gomide









Salve os dados obtidos em um arquivo CSV no seguinte formato:

**Título;Texto;Autor**


Exemplo:

**HISTÓRIAS DO MAR;Navio histórico vai afundar e se tornará maior recife artificial do mundo;Jorge de Souza**

Agora crie uma expressão regular para capturar o nome e a imagem de todos os colunistas da UOL.

 <b>Futebol pelo mundo</b> Futebol ou faculdade? Filho de Sylvinho une os dois nos EUA e sonha com MLS	 <b>Bruno Andrade</b> Mora em Lisboa desde 2015. Começou a carreira no LANCE! e depois virou c...	 <b>Allan Simon</b> YouTube anuncia Brasileiro na CazéTV em 2025 com todas as cotas vendidas	 <b>Rafael Reis</b> Vídeo de sexo faz árbitra e inspetor dos juizes serem suspensos na Turquia
 <b>Lei em Campo</b> STF tem hoje compromisso com esporte: julgar caso que define futuro da CBF	 <b>Pole Position</b> Colapinto diz que é mais difícil ser sul-americano na F1 e elogia Bortoleto	 <b>Saque e Voleio</b> Entrevista com Laura Pigossi: mais agressividade em busca de grande salto	 <b>Na Grade do MMA</b> Alex Poatan aniquilou rivais, mas ainda há desafio - e não é Jon Jones

<b>Imagem</b>	
<b>Nome</b>	<b>Yara Fantoni</b>

Como foi a ida de Memphis a jogo de várzea: 'Querida conhecer a favela'

Salve os dados obtidos em um arquivo CSV no seguinte formato:

**Nome ; Imagem**

Exemplo:

**Yara Fantoni ; https://conteudo.imguol.com.br/e6/2024/05/20/yara-fantoni-150-1716218084796\_150x150.jpg.webp**



## Laboratório

Curso Superior em Engenharia de Computação  
Disciplina de Linguagem Formais e Autômatos  
Professor Gomide

### Estudo de Caso B

Crie um programa que acesse o código-fonte do edital nº 10/2024 do Diário Oficial da União.

<https://www.in.gov.br/web/dou/-/edital-n-10/2024-589442586>

Crie uma ou várias expressões regulares que permita capturar todos os dados das pessoas que foram apresentadas na tabela presente no site. Os dados são: MATRÍCULA, NOME, ÓRGÃO e TIPO.

Na hipótese de moléstia grave ou de impossibilidade de locomoção que exija permanência domiciliar, o beneficiário ou o seu representante legal ou voluntário poderá solicitar à Unidade de Gestão de Pessoas o agendamento de visita técnica pela página <https://encurtador.com.br/nV1i7>, mediante apresentação de atestado ou laudo que comprove a impossibilidade do comparecimento para fins de prova de vida.

Nos casos em que a representação legal for exercida por tutor ou curador, a comprovação de vida deverá ser realizada exclusivamente na Unidade de Gestão de Pessoas do órgão de vinculação do beneficiário. O tutor ou curador deverá comparecer acompanhado do beneficiário, sendo indispensável a apresentação de original e cópia simples do termo de sentença judicial que o nomeou; Cadastro de Pessoa Física (CPF) do beneficiário; e documento oficial de identificação original com foto do beneficiário ou a sua Certidão de Nascimento, caso o beneficiário seja menor de dezoito anos:

Para mais informações, acessar a página [www.gov.br/servidor/decipex](http://www.gov.br/servidor/decipex) ou contatar a Central Decipex 0800 978 9004.

MATRÍCULA	NOME	ÓRGÃO	TIPO
134442	ADALCYR CUNHA DE SOUZA	40806	Aposentado
710188	ADALIA LUCIA MOURA SIQUEIRA	40806	Aposentado
104359	ADALVA ALICE DA CUNHA	40806	Aposentado
5837260	ADELIA CHAVES	40806	Beneficiário
3798836	ADELINA SOARES	40806	Beneficiário
2066457	ADEMAR SALVALAGGIO	40806	Aposentado
6897851	ADERITA GOMES PINHEIRO	40806	Beneficiário
3935833	ADILEUZA SAMPAIO TOSTA	40806	Beneficiário
1366014	ADONIAS DA SILVA COSTA	40806	Aposentado
1272679	AFONSO CELSO MESQUITA PEREIRA DE LIMA	40806	Aposentado
792099	AINA JOAQUINA IMBIRIBA GONCALVES	40806	Aposentado
6131522	ALAIDE BARBOSA VELOSO	40806	Beneficiário
6354122	ALAIDE WESTARP	40806	Beneficiário
5866383	ALBA SILVA ARAUJO	40806	Beneficiário
4583388	ALDA DOS SANTOS CHAVES	40805	Beneficiário
3372499	ALDA MARIA BORGES CUNHA	40806	Aposentado
2741288	ALDAI BASTOS FERNANDES	40806	Beneficiário
135315	ALDEMIR BARROS DO NASCIMENTO	40806	Aposentado
3386465	ALDENORA GIL BRITO	40806	Beneficiário
6845550	ALFONSO RODRIGUES DE LIMA	40806	Beneficiário



## Laboratório

Curso Superior em Engenharia de Computação  
Disciplina de Linguagem Formais e Autômatos  
Professor Gomide

Salve os dados obtidos em um arquivo CSV no seguinte formato:

**MATRÍCULA ; NOME ; ORGÃO ; TIPO**

Exemplo:

**2066457 ; ADEMAR SALVALAGGIO ; 40806 ; Aposentado**

Para finalizar, deseja-se extrair uma lista de todos: os gestores cape, telefones, CEPs, cidades e estados presentes nas tabelas.

252754	ZIMAR LAGES DE SOUZA	40806	Aposentado
6908462	ZOE MIRANDA FERREIRA	40806	Beneficiário
1665049	ZUHEYLA MARIA DE OLIVEIRA LIMA	40806	Beneficiário
4310411	ZULEIDE DA SILVA AGUIAR	40806	Beneficiário
1612136	ZULMIRA DE SOUZA BATISTA FERNANDES	40806	Aposentado
701347	ZULMIRA LIMA COSTA DA SILVA	40806	Aposentado

<b>UNIDADES DE GESTÃO DE PESSOAS</b>		<b>Telefone</b>
<b>Gestor</b>	Acre Central de Atendimento de Pessoal do Acre - CAPE/AC - Gestor da CAPE: <u>Geraldo Henrique Cavalcante de Carvalho</u> - Tel: <u>(68) 2106-8662</u>	Alagoas Central de Atendimento de Pessoal de Alagoas - CAPE/AL Gestora da CAPE: Maria Goret de França Rodrigues Tel: (82) 3311-2738 Praça Dom Pedro II, nº 16 - Bairro Centro CEP: 57020-130 Maceió/AL
<b>CEP</b>	Rua Benjamin Constant, nº 1.088 Bairro Cadeia Velha CEP: <u>69900-064</u> <u>Rio Branco</u> <u>AC</u>	
<b>Cidade</b>		<b>UF</b>
Amapá - Central de Atendimento de Pessoal do Amapá - CAPE/AP Gestora da CAPE: Celina de Souza Tel: (96) 3198-2713 Avenida Iracema Carvão Nunes, nº 625 - Bairro Central CEP: 68900-099 Macapá/AP		Amazonas - Central de Atendimento de Pessoal do Amazonas - CAPE/AM Gestora da CAPE: Vilani Ribeiro dos Santos Tel: (92) 3133-9086 Avenida Governador Danilo Areosa, nº 1530, Bloco B, 2º Andar, Quadrante 1 - Distrito Industrial CEP: 69075-351 Manaus/AM
Bahia - Central de Atendimento de Pessoal da Bahia - CAPE/BA Gestor da CAPE: Gustavo Bruni Teixeira Tel: (71) 3254-5265 Avenida Jequitaia, s/n, Edifício Sede do ME, Térreo - Bairro Comércio CEP: 40015-902 Salvador/ BA		Ceará - Central de Atendimento de Pessoal do Ceará - CAPE/CE Gestora da CAPE: Eliza Maria Jorge Vieira Tel: (85) 3878-3625 Rua Barão de Aracati, nº 909, 6º andar - Meireles CEP: 60115-080 Fortaleza/CE

Salve os dados obtidos em um arquivo CSV no seguinte formato:

**Gestor ; Telefone ; CEP ; Cidade ; UF**



## **Laboratório**

Curso Superior em Engenharia de Computação  
Disciplina de Linguagem Formais e Autômatos  
Professor Gomide

Exemplo:

```
Gustavo Bruni Teixeira; (71) 3254-5265; 40015-902; Salvador; BA
```

Dicas de Python:

- Para pegar um código-fonte de um site em Python use a biblioteca requests

```
pip install requests
```

- Para gerar o csv de forma bem simples, armazene os dados em um DataFrame. Esse objeto está presente na biblioteca Pandas

```
pip install pandas
```

- Ao capturar o código-fonte, guarde em uma variável e armazene esse conteúdo em um arquivo de texto. Depois trabalhe os estudos de casos em cima dos arquivos de texto gerados. Essa ação lhe poupará **MUITO TEMPO** de processamento. Aqui vai um código simples para gerar um arquivo de texto:

```
try:
```

```
    with open(nome_arquivo, "w", encoding="utf-8") as arquivo:
```

```
        arquivo.write(codigo_fonte)
```

```
    print(f"Arquivo '{nome_arquivo}' salvo.")
```

```
except IOError as e:
```

```
    print(f"Erro ao salvar o arquivo: {e}")
```