

# cebrap.lab - Raspagem de Dados com R

Thiago Meireles

Nic.br | Cetic.br

23 de outubro de 2023

# Apresentação

- O objetivo do curso é introduzir a raspagem de dados na Internet utilizando a linguagem R.
- A partir de uma abordagem focada na prática, diferentes estratégias de capturas de dados em páginas de internet serão apresentadas.
- A quantidade de informação útil (organizada ou não) em sites torna fundamental a utilização de ferramentas que facilitem sua captura e compilação.

# Requisitos

- A participação no curso requer uma exposição prévia à linguagem R e ao ambiente de trabalho do RStudio.
- Também foi indicada a realização de tutoriais de preparação antes do início das aulas.

# Objetivos

- Os participantes, ao fim do curso, serão capazes de:
  - 1 Coletar dados de sites de estrutura definida;
  - 2 Automatizar a coleta de conteúdo em sites com a mesma estrutura;
  - 3 Estruturar o conteúdo dos sites em bases de dados, com textos *limpos* e organizados;
  - 4 Entender as diferenças na coleta em páginas dinâmicas.

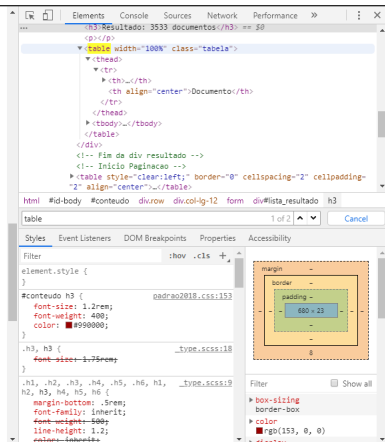
# Roteiro e Calendário

- 23/10/2023 - O básico da raspagem de dados
- 24/10/2023 - Aplicação da raspagem de dados em pesquisas
- 25/10/2023 - Automatizando a raspagem de dados e manipulando o conteúdo
- 26/10/2023 - Desafios em raspagem de dados
- 27/10/2023 - Páginas Dinâmicas

# Tabelas

Coronel Telhada	<b>Indicação 3209/2019, de 25/09/2019</b> Indica ao Sr. Governador que insira o inciso XVII e a alínea "a" ao artigo 78 da Lei 10.261/68, para dispor sobre a licença paternidade para viúvo.
Leticia Aguiar	<b>Moção 144/2019, de 20/09/2019</b> Apela ao Sr. Presidente do Senado Federal, Davi Alcolumbre, para que os guardas civis municipais e metropolitanos tenham tratamento igualitário aos demais integrantes das forças de segurança pública, garantindo a eles, dessa forma, justiça social na Reforma da Previdência.
Rodrigo Gambale	<b>Projeto de lei 1059/2019, de 19/09/2019</b> Obriga os órgãos públicos competentes a manter campanha publicitária permanente de orientação para os motociclistas no âmbito do Estado.
Henri Ozi Cukier	<b>Parecer 1103/2019, de 18/09/2019 - Projeto de lei 512/2019</b> Dispõe sobre o funcionamento de Gabinetes Optométricos de profissionais habilitados para o atendimento à saúde visual primária na rede privada do Estado de São Paulo.
Daniel José	<b>Emenda de Pauta 603, de 18/09/2019 - Projeto de lei 924/2019</b> MODIFICA O INDICADOR "TEMPO MÉDIO DE CONCESSÃO DE APOSENTADORIA (DIA)" DO PROGRAMA 2021 - GESTÃO DOS REGIMES PRÓPRIOS DE PREVIDÊNCIA ESTADUAL, na seguinte conformidade:
Daniel José	<b>Emenda de Pauta 599, de 18/09/2019 - Projeto de lei 924/2019</b> MODIFICA O INDICADOR DE PRODUTO DO PROGRAMA 2021 - GESTÃO DOS REGIMES PRÓPRIOS DE PREVIDÊNCIA ESTADUAL, na seguinte conformidade:
Daniel José	<b>Emenda de Pauta 598, de 18/09/2019 - Projeto de lei 924/2019</b> MODIFICA O INDICADOR DE RESULTADO DO PROGRAMA 2021 - GESTÃO DOS REGIMES PRÓPRIOS DE PREVIDÊNCIA ESTADUAL, na seguinte conformidade:

table.tabela 680 x 1682



# Nós (Nodes)

Coronel Tethada

## Indicação 3209/2019, de 25/09/2019

Indica ao Sr. Governador que insira o inciso XVII e a alínea "a" ao artigo 78 da Lei 10.261/68, para dispor sobre a licença paternidade para viúvo.

Letícia Aguiar

## Moção 144/2019, de 20/09/2019

Apela ao Sr. Presidente do Senado Federal, Davi Alcolumbre, para que os guardas civis municipais e metropolitanos tenham tratamento igualitário aos demais integrantes das forças de segurança pública, garantindo a eles, dessa forma, justiça social na Reforma

da Democracia.

Rodrigo Gambale

## Projeto de lei 1059/2019, de 19/09/2019

Obriga os órgãos públicos competentes a manter campanha publicitária permanente de orientação para os motociclistas no âmbito do Estado.

Henri Ozi Cukier

## Parecer 1103/2019, de 18/09/2019 - Projeto de lei 512/2019

Dispõe sobre o funcionamento de Gabinetes Optométricos de profissionais habilitados para o atendimento à saúde visual primária na rede privada do Estado de São Paulo.

Daniel José

## Emenda de Pauta 603, de 18/09/2019 - Projeto de lei 924/2019

MODIFICA O INDICADOR "TEMPO MÉDIO DE CONCESSÃO DE APOSENTADORIA (DIA)" DO PROGRAMA 2021 E GESTÃO DOS REGIMES PRÓPRIOS DE PREVIDÊNCIA ESTADUAL na seguinte conformidade:

Daniel José

## Emenda de Pauta 599, de 18/09/2019 - Projeto de lei 924/2019

MODIFICA O INDICADOR DE PRODUTO DO PROGRAMA 2021 E GESTÃO DOS REGIMES PRÓPRIOS DE PREVIDÊNCIA ESTADUAL na seguinte conformidade:

Daniel José

## Emenda de Pauta 598, de 18/09/2019 - Projeto de lei 924/2019

MODIFICA O INDICADOR DE RESULTADO DO PROGRAMA 2021 - GESTÃO DOS REGIMES PRÓPRIOS DE PREVIDÊNCIA ESTADUAL na seguinte conformidade:

The screenshot shows the browser's developer tools. The DOM tree on the left highlights a `strong` element within a table. The Styles pane on the right shows the default styles for the `strong` element, including `color: #000000`, `background-color: transparent`, and `text-decoration: none`. The background shows a snippet of HTML code for a table row with a `strong` element.

# Obtendo os caminhos do nó

Coronel Telhada	<b>Indicação 3209/2019, de 25/09/2019</b> Indica ao Sr. Governador que insira o inciso XVII e a alínea "a" ao artigo 78 da Lei 10.261/68, para dispor sobre a licença paternidade para viúvo.
Leticia Aguiar	<b>Moção 144/2019, de 20/09/2019</b> Apela ao Sr. Presidente do Senado Federal, Davi Alcolumbre, para que os guardas civis municipais e metropolitanos tenham tratamento igualitário aos demais integrantes das forças de segurança pública, garantindo a eles, dessa forma, justiça social na Reforma
Rodrigo Gambale	<b>Projeto de lei 1059/2019, de 19/09/2019</b> Obriga os órgãos públicos competentes a manter campanha publicitária permanente de orientação para os motociclistas no âmbito do Estado.
Henri Ozi Cukier	<b>Parecer 1103/2019, de 18/09/2019 - Projeto de lei 512/2019</b> Dispõe sobre o funcionamento de Gabinetes Optométricos de profissionais habilitados para o atendimento à saúde visual primária na rede privada do Estado de São Paulo.
Daniel José	<b>Emenda de Pauta 603, de 18/09/2019 - Projeto de lei 924/2019</b> MODIFICA O INDICADOR "TEMPO MÉDIO DE CONCESSÃO DE APOSENTADORIA (DIA)" DO PROGRAMA 2021 E GESTÃO DOS REGIMES PRÓPRIOS DE PREVIDÊNCIA ESTADUAL, na seguinte conformidade:
Daniel José	<b>Emenda de Pauta 599, de 18/09/2019 - Projeto de lei 924/2019</b> MODIFICA O INDICADOR DE PRODUTO DO PROGRAMA 2021 E GESTÃO DOS REGIMES PRÓPRIOS DE PREVIDÊNCIA ESTADUAL, na seguinte conformidade:
Daniel José	<b>Emenda de Pauta 598, de 18/09/2019 - Projeto de lei 924/2019</b> MODIFICA O INDICADOR DE RESULTADO DO PROGRAMA 2021 - GESTÃO DOS REGIMES PRÓPRIOS DE PREVIDÊNCIA ESTADUAL, na seguinte conformidade:

