

cebrap.lab - Captura de Dados com R

Thiago Meireles

Universidade de São Paulo — Departamento de Ciência Política
Cetic.br — Nic.br

23 de maio de 2022

Apresentação

- O objetivo do curso é introduzir a captura 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 mais simples, como jornais e legislativos brasileiros;
 - 2 Realizar tarefas relacionadas a mineração de texto a partir de diferentes abordagens
 - 3 Produzir gráficos e grafos mais simples a partir dos dados coletados
 - 4 Entender e aplicar conceitos básicos de *text mining*.

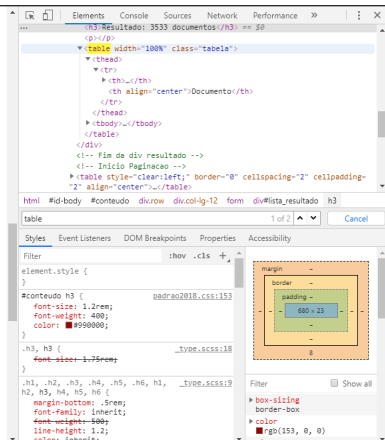
Roteiro e Calendário

- 07/06/2021 - O básico da raspagem de dados
- 08/06/2021 - Desafios de raspagem de dados
- 09/06/2021 - Introdução à manipulação de textos como dados
- 10/06/2021 - A pesquisa com *Text mining*
- 11/06/2021 - *Text mining* em R

Tabelas

Coronel Telhada	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.
Leticia 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 Previdência.
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 - 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 - 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:

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 a web browser's developer tools. The DOM tree on the left shows a table with a row containing a strong element. The Styles pane on the right shows the default styles for the strong element, including color, background-color, and text-decoration.

```

... #conteudo div div form #lista_resultado table tbody tr td a strong
table
  1 of 2
  Cancel
  Styles Event Listeners DOM Breakpoints Properties Accessibility
  Filter
  element.style {
    color: #000000;
  }
  a {
    color: #000000;
  }
  a {
    color: #0000ff;
    text-decoration: none;
    background-color: transparent;
    text-decoration: none;
  }
  a {
    font-family: "gisha";
  }
  .:after .:before {
    rehint: 0.000000;
  }
  
```

The Styles pane shows the default styles for the strong element, including color, background-color, and text-decoration.

Obtendo os caminhos do nó

Coronel Telhada	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.
Leticia 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
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:

