

Carregando e armazenando dados em vários formatos de arquivos

Trabalhando com XML

Carregando e armazenando dados

Trabalhando com XML

O XML (Extensible Markup Language, ou Linguagem de Marcação Extensível) é um formato de dados estruturados que aceita dados hierárquicos e aninhados com metadados.

É um dos subtipos da SGML (acrônimo de Standard Generalized Markup Language ou Linguagem Padronizada de Marcação Genérica) capaz de descrever diversos tipos de dados. Seu propósito principal é a facilidade de compartilhamento de informações por intermédio da internet, sendo de fácil entendimento para humanos e sistemas.

Carregando e armazenando dados

Trabalhando com XML

É semelhante ao HTML, porém, no HTML temos tags pré-definidas, já no XML, as tags são criadas livremente. Veja um exemplo.

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <receita nome="pão" tempo_de_preparo="5 minutos" tempo_de_cozimento="1 hora">
3   <titulo>Pão simples</titulo>
4   <ingredientes>
5     <ingrediente quantidade="3" unidade="xícaras">Farinha de Trigo</ingrediente>
6     <ingrediente quantidade="7" unidade="gramas">Fermento</ingrediente>
7     <ingrediente quantidade="1.5" unidade="xícaras" estado="morna">Água</ingrediente>
8     <ingrediente quantidade="1" unidade="colheres de chá">Sal</ingrediente>
9   </ingredientes>
10  <instrucoes>
11    <passo>Misture todos os ingredientes, e dissolva bem.</passo>
12    <passo>Cubra com um pano e deixe por uma hora em um local morno.</passo>
13    <passo>Misture novamente, coloque numa bandeja e asse num forno.</passo>
14  </instrucoes>
15 </receita>
```

Carregando e armazenando dados

Trabalhando com XML

“receita” é o nome principal para o seu documento. Note que a semelhança entre XML e HTML é grande, na 1ª linha abrimos a *tag* receita e na última linha a fechamos, como em [HTML](#), assim se estendendo por todo o exemplo.

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <receita nome="pão" tempo_de_preparo="5 minutos" tempo_de_cozimento="1 hora">
3   <titulo>Pão simples</titulo>
4   <ingredientes>
5     <ingrediente quantidade="3" unidade="xícaras">Farinha de Trigo</ingrediente>
6     <ingrediente quantidade="7" unidade="gramas">Fermento</ingrediente>
7     <ingrediente quantidade="1.5" unidade="xícaras" estado="morna">Água</ingrediente>
8     <ingrediente quantidade="1" unidade="colheres de chá">Sal</ingrediente>
9   </ingredientes>
10  <instrucoes>
11    <passo>Misture todos os ingredientes, e dissolva bem.</passo>
12    <passo>Cubra com um pano e deixe por uma hora em um local morno.</passo>
13    <passo>Misture novamente, coloque numa bandeja e asse num forno.</passo>
14  </instrucoes>
15 </receita>
```

Carregando e armazenando dados

Trabalhando com XML

Veja conteúdo do arquivo de exemplo aberto em um navegador.

```
<?xml version="1.0" encoding="UTF-8"?>
- <receita tempo_de_cozimento="1 hora" tempo_de_preparo="5 minutos" nome="pão">
  <titulo>Pão simples</titulo>
  - <ingredientes>
    <ingrediente unidade="xícaras" quantidade="3">Farinha de Trigo</ingrediente>
    <ingrediente unidade="gramas" quantidade="7">Fermento</ingrediente>
    <ingrediente unidade="xícaras" quantidade="1.5" estado="morna">Água</ingrediente>
    <ingrediente unidade="colheres de chá" quantidade="1">Sal</ingrediente>
  </ingredientes>
  - <instrucoes>
    <passo>Misture todos os ingredientes, e dissolva bem.</passo>
    <passo>Cubra com um pano e deixe por uma hora em um local morno.</passo>
    <passo>Misture novamente, coloque numa bandeja e asse num forno.</passo>
  </instrucoes>
</receita>
```

Carregando e armazenando dados

Trabalhando com XML

Para trabalhar com XML, vamos utilizar a biblioteca padrão ElementTree.

Primeiro vamos importar essa biblioteca juntamente com o Pandas.

```
import xml.etree.ElementTree as ET  
import pandas as pd
```

Carregando e armazenando dados

Trabalhando com XML

Vamos agora fazer o parse do arquivo receita.xml para uma variável chamada receitas.

```
receitas = ET.parse("receita.xml")
```

Vamos carregar o nó raiz para dentro da variável root.

```
root = receitas.getroot()
```

Carregando e armazenando dados

Trabalhando com XML

Utilize o `root.tag` para ver o nome da tag ("receita").

```
print(root.tag)
```

Utilize o `root.attrib` para ver os atributos da tag ("receita").

```
print(root.attrib)
```


Carregando e armazenando dados

Trabalhando com XML

Vamos imprimir o nome de cada subelemento de root.

```
for child in root:  
    print(child.tag)
```

Percorrendo todos elementos do arquivo usando compreensão de lista (List Comprehension).

```
print([elem.tag for elem in root.iter()])
```

Carregando e armazenando dados

Trabalhando com XML

Vamos criar uma lista vazia chamada dados.

```
dados = []
```

Carregando e armazenando dados

Trabalhando com XML

Vamos percorrer cada nó "ingrediente" usando `root.iter('ingrediente')` e vamos utilizar sua tag, texto e atributos.

```
for ingrediente in root.iter('ingrediente'):
    dado = {}
    dado[ingrediente.tag] = ingrediente.text
    dado['atributos'] = ingrediente.attrib
    dados.append(dado)
```

Carregando e armazenando dados

Trabalhando com XML

Agora vamos carregar os dados para um DataFrame e vamos imprimir seu conteúdo.

```
df = pd.DataFrame(dados)
print(f"df: {df}")
```

FIM