



# Writing and Reading Data Files

Catarina Oliveira

**DCT** DEPARTAMENTO CIÊNCIA  
E TECNOLOGIA

## CONTENT

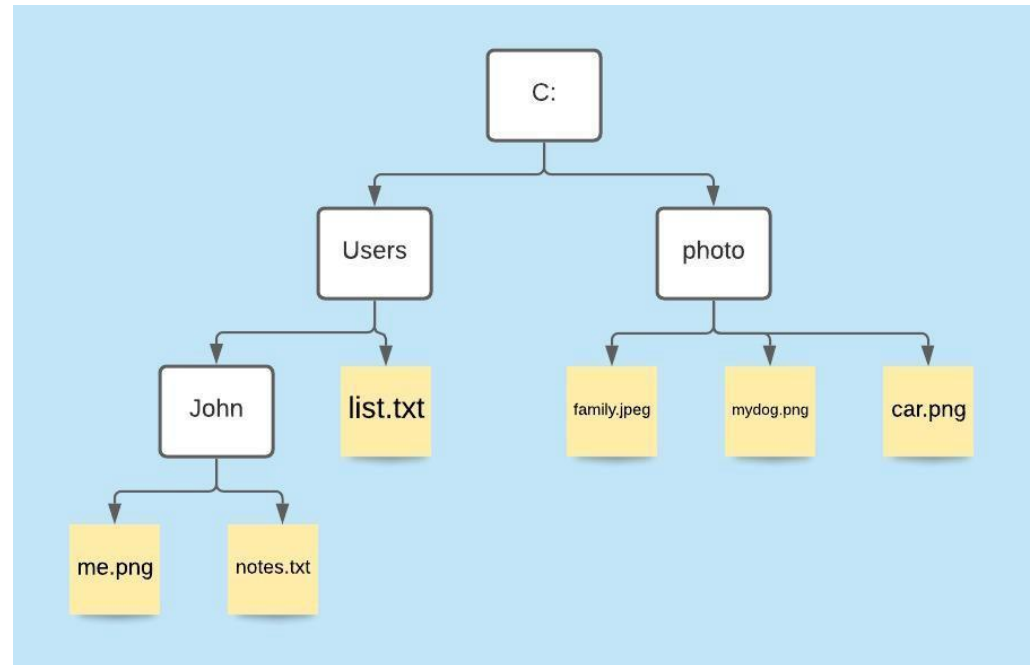
1. file system
2. open/close files
3. file reading
4. file writing
5. Special file types
  1. Read/Write CSV
  2. Read/Write JSON
  3. Read/Write XML
6. Read Web Pages

## file system

- identified by names
  - Eg: **car.png**
- Organized hierarchically in folders
  - Ex (Windows): **C:/ photo /car.png**
  - Ex (Linux): **/ photo /car.png**

What would be the path to the **notes.txt** file?

- Windows
- Linux



## open/close files

- In Python, files can be opened ( `open` ) in several ways:
  - `r`: read
  - `w`: write
  - `a`: add
  - `x`: create
- At the end of use they must be closed ( `close` )
- Example: Open `file.txt` in read mode and print the file name:

```
ficheiro = open("test.txt", "r")  
print(ficheiro.name)  
ficheiro.close()
```

```
with open("test.txt", "r") as ficheiro:  
    print(ficheiro.name)
```

## file reading

```
with open("test.txt", "r") as ficheiro:  
    ficheiro.read() # lê o ficheiro completo  
    ficheiro.readline() # lê uma linha do ficheiro  
    ficheiro.readlines() # lê o ficheiro completo como uma lista de linhas  
    ficheiro.read(5) # lê 5 caracteres
```

```
with open("test.txt", "r") as ficheiro:  
    for linha in ficheiro:  
        (...)  
        # vai lendo uma linha de cada vez  
        # para não ficar o ficheiro todo em memória
```

## file writing

- Open a file and write a string (overrides if existing)

```
with open("test.txt","w") as ficheiro:  
    ficheiro.write("Era\n")
```

- Open a file and add two strings one at a time

```
with open("test.txt","a") as ficheiro:  
    ficheiro.write("uma\n")  
    ficheiro.write("vez\n")
```

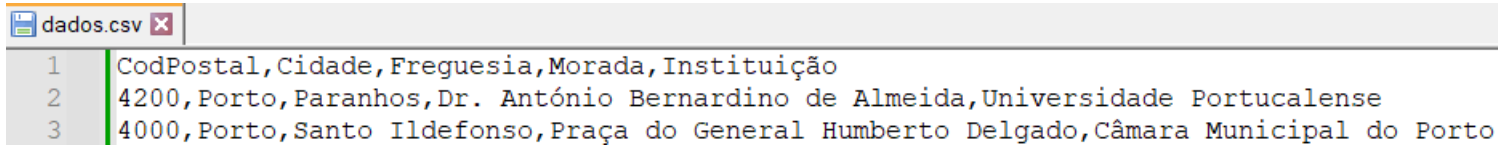
- Open a file and add two lines from a list of strings

```
with open("test.txt","a") as ficheiro:  
    ficheiro.writelines(["em\n","Marte\n"])
```

## Special file types

- Python has specific modules for processing files:
  - csv
  - JSON
  - XML
- ... and also for processing:
  - Web Pages
- ... between others

## Read CSV (to a dictionary)



```
dados.csv x
1 CodPostal,Cidade,Freguesia,Morada,Instituição
2 4200,Porto,Paranhos,Dr. António Bernardino de Almeida,Universidade Portucalense
3 4000,Porto,Santo Ildefonso,Praça do General Humberto Delgado,Câmara Municipal do Porto
```

```
import csv

dadosObtidos = []
with open("dados.csv", "r", encoding="utf8") as ficheiro:
    dicionario = csv.DictReader(ficheiro)
    for linha in dicionario:
        dadosObtidos.append(linha)

print(dadosObtidos)
```

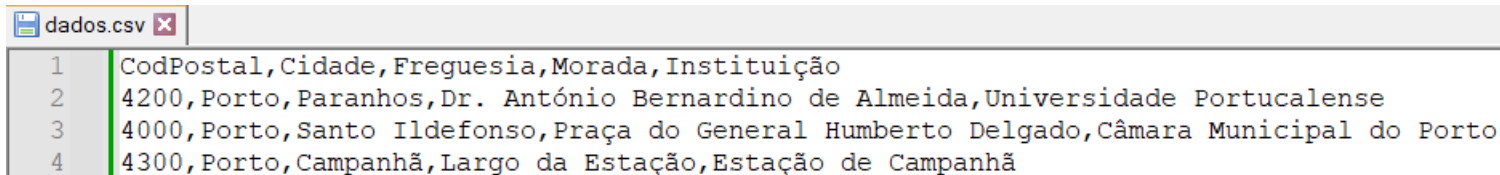
```
[
  {
    'CodPostal': '4200',
    'Cidade': 'Porto',
    'Freguesia': 'Paranhos',
    'Morada': 'Dr. António Bernardino de Almeida',
    'Instituição': 'Universidade Portucalense'
  },
  {
    'CodPostal': '4000',
    'Cidade': 'Porto',
    'Freguesia': 'Santo Ildefonso',
    'Morada': 'Praça do General Humberto Delgado',
    'Instituição': 'Câmara Municipal do Porto'
  }
]
```



## Write CSV (from a dictionary)

```
estacaoCampanha = {  
    'CodPostal': '4300',  
    'Cidade': 'Porto',  
    'Freguesia': 'Campanhã',  
    'Morada': 'Largo da Estação',  
    'Instituição': 'Estação de Campanhã'  
}
```

```
import csv  
  
with open("dados.csv", "a", encoding="utf8", newline='\n') as ficheiro:  
    writer = csv.DictWriter(ficheiro, fieldnames=estacaoCampanha.keys())  
    writer.writerow(estacaoCampanha)
```



The screenshot shows a text editor window titled 'dados.csv'. It contains four lines of CSV data, each preceded by a line number (1-4). The data represents the 'estacaoCampanha' dictionary entries, with fields separated by commas.

	CodPostal	Cidade	Freguesia	Morada	Instituição
1	CodPostal	Cidade	Freguesia	Morada	Instituição
2	4200	Porto	Paranhos	Dr. António Bernardino de Almeida	Universidade Portucalense
3	4000	Porto	Santo Ildefonso	Praça do General Humberto Delgado	Câmara Municipal do Porto
4	4300	Porto	Campanhã	Largo da Estação	Estação de Campanhã

## Read JSON (to a dictionary)

```

dados.json x
1 {
2   "instituições": [
3     {
4       "CodPostal": 4200,
5       "Cidade": "Porto",
6       "Freguesia": "Paranhos",
7       "Morada": "Rua Dr. António Bernardino de Almeida",
8       "Instituição": "Universidade Portucalense"
9     },
10    {
11      "CodPostal": 4000,
12      "Cidade": "Porto",
13      "Freguesia": "Santo Ildefonso",
14      "Morada": "Praça do General Humberto Delgado",
15      "Instituição": "Câmara Municipal do Porto"
16    }
17  ]
18 }

```

```

import json

with open("dados.json", "r", encoding="utf8") as ficheiro:
    dicionario = json.load(ficheiro)

print(dicionario)

```

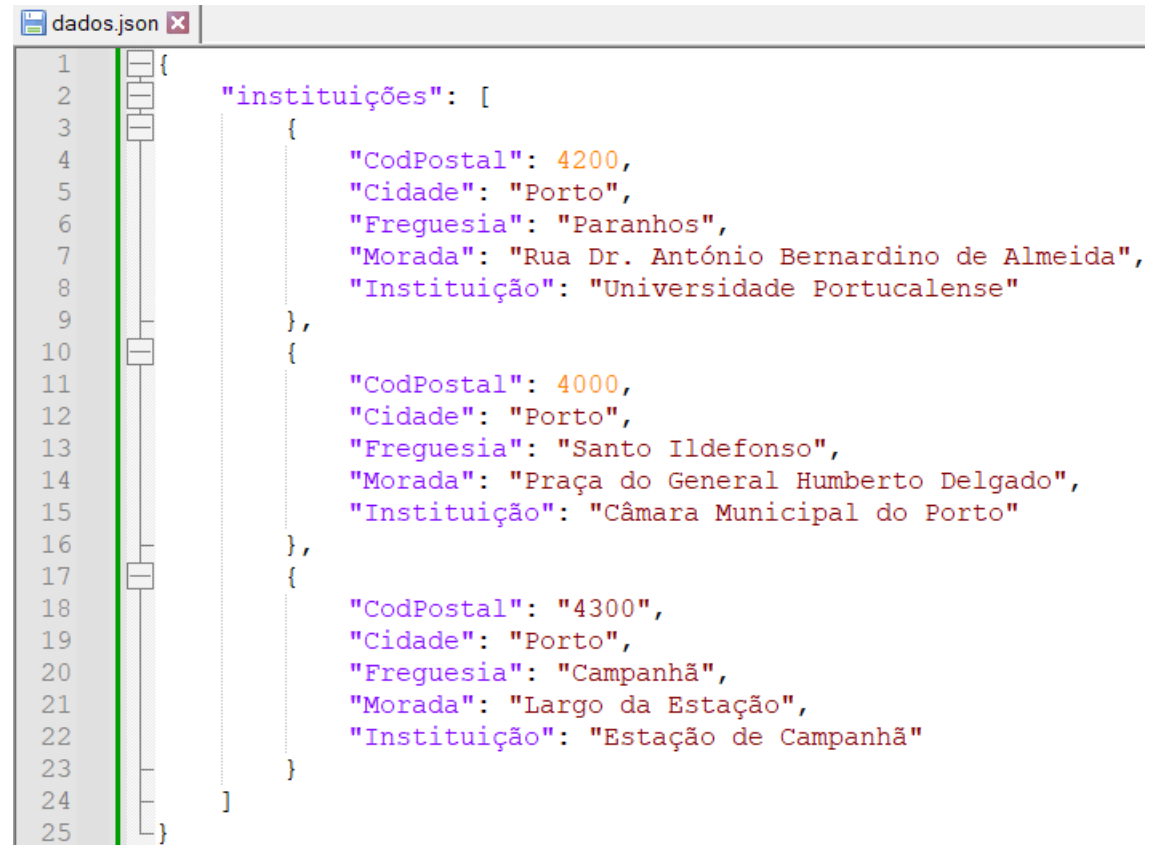
```

{
  "instituições": [
    {
      "CodPostal": 4200,
      "Cidade": "Porto",
      "Freguesia": "Paranhos",
      "Morada": "Rua Dr. António Bernardino de Almeida",
      "Instituição": "Universidade Portucalense"
    },
    {
      "CodPostal": 4000,
      "Cidade": "Porto",
      "Freguesia": "Santo Ildefonso",
      "Morada": "Praça do General Humberto Delgado",
      "Instituição": "Câmara Municipal do Porto"
    }
  ]
}

```

## Write JSON (from a dictionary)

```
estacaoCampanha = {
    'CodPostal': '4300',
    'Cidade': 'Porto',
    'Freguesia': 'Campanhã',
    'Morada': 'Largo da Estação',
    'Instituição': 'Estação de Campanhã'
}
```



```
dados.json
1 {
2   "instituições": [
3     {
4       "CodPostal": 4200,
5       "Cidade": "Porto",
6       "Freguesia": "Paranhos",
7       "Morada": "Rua Dr. António Bernardino de Almeida",
8       "Instituição": "Universidade Portucalense"
9     },
10    {
11      "CodPostal": 4000,
12      "Cidade": "Porto",
13      "Freguesia": "Santo Ildefonso",
14      "Morada": "Praça do General Humberto Delgado",
15      "Instituição": "Câmara Municipal do Porto"
16    },
17    {
18      "CodPostal": "4300",
19      "Cidade": "Porto",
20      "Freguesia": "Campanhã",
21      "Morada": "Largo da Estação",
22      "Instituição": "Estação de Campanhã"
23    }
24  ]
25 }
```

After reading:

```
dicionario["instituições"].append(estacaoCampanha)
with open("dados.json", "w", encoding="utf8", newline='\n') as ficheiro:
    writer = json.dump(dicionario, ficheiro)
```

## Read XML (to a dictionary)

```

1 <dados>
2   <instituição>
3     <CodPostal>4200</CodPostal>
4     <Cidade>Porto</Cidade>
5     <Freguesia>Paranhos</Freguesia>
6     <Morada>Dr. António Bernardino de Almeida</Morada>
7     <Instituição>Universidade Portucalense</Instituição>
8   </instituição>
9   <instituição>
10    <CodPostal>4000</CodPostal>
11    <Cidade>Porto</Cidade>
12    <Freguesia>Santo Ildefonso</Freguesia>
13    <Morada>Praça do General Humberto Delgado</Morada>
14    <Instituição>Câmara Municipal do Porto</Instituição>
15  </instituição>
16 </dados>

```

```

import xmltodict

with open('dados.xml', 'r', encoding='utf-8') as ficheiro:
    my_xml = ficheiro.read()
    dicionario = xmltodict.parse(my_xml)

print(dicionario)

```

```

{
  'dados':
  {
    'instituição':
    [
      {
        'CodPostal': '4200',
        'Cidade': 'Porto',
        'Freguesia': 'Paranhos',
        'Morada': 'Dr. António Bernardino de Almeida',
        'Instituição': 'Universidade Portucalense'
      },
      {
        'CodPostal': '4000',
        'Cidade': 'Porto',
        'Freguesia': 'Santo Ildefonso',
        'Morada': 'Praça do General Humberto Delgado',
        'Instituição': 'Câmara Municipal do Porto'
      }
    ]
  }
}

```

## Write XML (from a dictionary)

```
estacaoCampanha = {
    'CodPostal': '4300',
    'Cidade': 'Porto',
    'Freguesia': 'Campanhã',
    'Morada': 'Largo da Estação',
    'Instituição': 'Estação de Campanhã'
}
```

After reading:

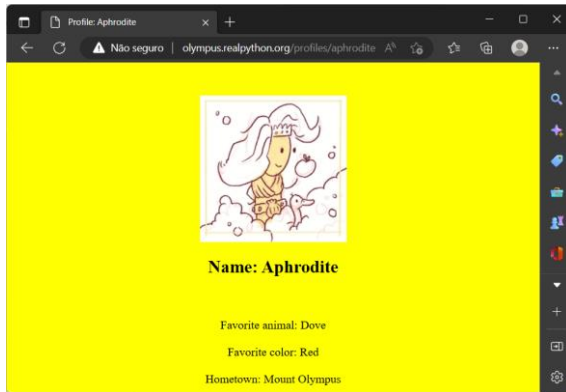
```
dicionario["dados"]["instituição"].append(estacaoCampanha)
xml_string = xmltodict.unparse(dicionario)

print(xml_string)

with open('dados.xml', 'w', encoding='utf-8') as ficheiro:
    ficheiro.write(xml_string)
```

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <dados>
3   <instituição>
4     <CodPostal>4200</CodPostal>
5     <Cidade>Porto</Cidade>
6     <Freguesia>Paranhos</Freguesia>
7     <Morada>Dr. António Bernardino de Almeida</Morada>
8     <Instituição>Universidade Portucalense</Instituição>
9   </instituição>
10  <instituição>
11    <CodPostal>4000</CodPostal>
12    <Cidade>Porto</Cidade>
13    <Freguesia>Santo Ildefonso</Freguesia>
14    <Morada>Praça do General Humberto Delgado</Morada>
15    <Instituição>Câmara Municipal do Porto</Instituição>
16  </instituição>
17  <instituição>
18    <CodPostal>4300</CodPostal>
19    <Cidade>Porto</Cidade>
20    <Freguesia>Campanhã</Freguesia>
21    <Morada>Largo da Estação</Morada>
22    <Instituição>Estação de Campanhã</Instituição>
23  </instituição>
24  <instituição>
25    <CodPostal>4300</CodPostal>
26    <Cidade>Porto</Cidade>
27    <Freguesia>Campanhã</Freguesia>
28    <Morada>Largo da Estação</Morada>
29    <Instituição>Estação de Campanhã</Instituição>
30  </instituição>
31 </dados>
```

## Read Web Pages



```
from urllib.request import urlopen
```

```
url = "http://olympus.realpython.org/profiles/aphrodite"
page = urlopen(url)
html_bytes = page.read()
html = html_bytes.decode("utf-8")
print(html)
```

```
<html>
  <head>
    <title>Profile: Aphrodite</title>
  </head>
  <body bgcolor="yellow">
    <center><br><br>
      
      <h2>Name: Aphrodite</h2><br><br>
      Favorite animal: Dove<br><br>
      Favorite color: Red<br><br>
      Hometown: Mount Olympus
    </center>
  </body>
</html>
```

```
<html>
<head>
<title>Profile: Aphrodite</title>
</head>
<body bgcolor="yellow">
<center>
<br><br>

<h2>Name: Aphrodite</h2>
<br><br>
Favorite animal: Dove
<br><br>
Favorite color: Red
<br><br>
Hometown: Mount Olympus
</center>
</body>
</html>
```



UNIVERSIDADE  
PORTUCALENSE

Do conhecimento à prática.