

# Extrair Vértices de Memorial Descritivo

O objetivo deste *notebook* é demonstrar a extração de coordenadas geodésicas referente a um memorial descritivo da Unidade de Manejo Florestal 4, localizada na Floresta Nacional de Amapá. O arquivo está disponível neste [link](#).

Utilizaremos o módulo `pdfminer.high_level` do pacote `pdfminer`.

```
In [1]: # Carregar o módulo extract_text do pacote pdfminer
from pdfminer.high_level import extract_text
```

## Ler como texto texto apenas as páginas 26, 27 e 28 do arquivo pdf

```
In [3]: # Ler páginas 26, 27 e 28
pdf_txt = extract_text('./data/ContratoConcesso012021_umfl_flonas_amapa.pdf', page_numbers=[26, 27, 28])
```

## Mostrar o conteúdo das páginas selecionadas

```
In [10]: # Mostrar o conteúdo das páginas
print(pdf_txt)
```

Unidade de Manejo Florestal (UMF) IV

Área Plana: 39.356,25 ha

Perímetro: 125.469,24 m

Municípios: Ferreira Gomes/AP

O limite da UMF IV foi definido seguindo orientações da equipe da Gerência Executiva de Monitoramento e Auditoria Florestal (GEMAF), utilizando como referência a hidrografia da Base Continua 1:250.000 (Bc250) versão 2017, da Coordenação de Cartografia da Diretoria de Geociências (DGC) do IBGE, Gerência de Bases Contínuas (GBC) e ainda utilizando imagens de Modelo digital de Superfície SRTM de 30m da USGS. O Perímetro forma um polígono irregular de 22 vértices. Inicia-se a descrição deste perímetro no vértice P-01, de coordenadas N 140.443,12m e E 410.769,20m, localizado na margem esquerda do Rio Santo Antônio; daí, segue a montante pelos meandros do igarapé sem nome por aproximadamente e distância de 4.991,2m até o ponto P-02, de coordenadas N 139.245,91m e E 415.407,42m, localizado na margem esquerda de igarapé sem nome; daí, segue em linha seca com o azimute plano 148°43'48" e distância de 2.323,2m até o ponto P-03, de coordenadas N 137.260,28m e E 416.613,41m, daí, segue em linha seca com o azimute plano 67°56'24" e distância de 2.322,3m até o ponto P-04, de coordenadas N 138.132,31m e E 418.765,81m, localizado na margem direita de igarapé sem nome; daí, segue a jusante pelos meandros do igarapé sem nome por aproximadamente e distância de 9.479,8m até o ponto P-05, de coordenadas N 130.563,09m e E 423.378,13m, localizado na confluência de dois igarapés sem nome; daí, segue a montante pelos meandros do igarapé sem nome por aproximadamente e distância de 496,0m até o ponto P-06, de coordenadas N 131.013,35m e E 423.560,87m, localizado em uma bifurcação para dois igarapés sem nome; daí, segue pelos meandros do igarapé da direita, a montante, por aproximadamente e distância de 5.399,4m até o ponto P-07, de coordenadas N 130.950,90m e E 428.236,94m, localizado na margem esquerda de igarapé sem nome; daí, segue em linha seca com o azimute plano 159°36'36" e distância de 164,3m até o ponto P-08, de coordenadas N 130.796,93m e E 428.294,17m, daí, segue em linha seca com o azimute plano 197°5'60" e distância de 4.951,8m até o ponto P-09, de coordenadas N 126.064,02m e E 426.838,05m, daí, segue em linha seca com o azimute plano 253°18'36" e distância de 1.995,7m até o ponto P-10, de coordenadas N 125.490,99m e E 424.926,37m, daí, segue em linha seca com o azimute plano 222°45'0" e distância de 2.702,8m até o ponto P-11, de coordenadas N 123.506,24m e E 423.091,73m, daí, segue em linha seca com o azimute plano 301°11'60" e distância de 1.210,2m até o ponto P-12, de coordenadas N 124.133,18m e E 422.056,57m, daí, segue em linha seca com o azimute plano 239°51'36" e distância de 2.790,9m até o ponto P-13, de coordenadas N 122.731,73m e E 419.643,04m, daí, segue em linha seca com o azimute plano 270°14'24" e distância de 2.137,6m até o ponto P-14, de coordenadas N 122.740,51m e E 417.505,46m, daí, segue em linha seca com o azimute plano 238°5'60" e distância de 858,6m até o ponto P-15, de coordenadas N 122.286,80m e E 416.776,58m, daí, segue

Contrato de Concessão Florestal nº 01/2021 – Anexo 1 – Página 13 de 15

em linha seca com o azimute plano 215°50'60" e distância de 9.552,3m até o ponto P-16, de coordenadas N 114.544,32m e E 411.181,91m, daí, segue em linha seca com o azimute plano 158°25'12" e distância de 2.907,0m até o ponto P-17, de coordenadas N 111.841,13m e E 412.251,32m, daí, segue em linha seca com o azimute plano 176°13'12" e distância de 2.037,1m até o ponto P-18, de coordenadas N 109.808,42m e E 412.385,57m, localizada na margem esquerda do Rio Araguari; daí, segue a jusante pelos meandros do Rio Araguari por aproximadamente e distância de 25.900,56 até o ponto P-19, de coordenadas N 115.818,01m e E 399.342,22m, localizado na margem esquerda do Rio Araguari; daí, segue em linha seca com o azimute plano 34°26'24" e distância de 2.317,7m até o ponto P-20, de coordenadas N 117.729,48m e E 400.652,90m, daí, segue em linha seca com o azimute plano 2°35'24" e distância de 2.663,0m até o ponto P-21, de coordenadas N 120.389,80m e E 400.773,23m, daí, segue em linha seca com o azimute plano 296°48'0" e distância de 628,2m até o ponto P-22, de coordenadas N 120.672,98m e E 400.212,53m, localizado na margem esquerda do Rio Santo Antônio; daí, segue a jusante pelos meandros do Rio Santo Antônio por aproximadamente e distância de 30.171,8m retornando-se ao ponto P-01 início desta descrição, fechando o polígono.

Fica excluída, do perímetro acima descrito, área encravada, cuja descrição do perímetro se inicia no P-23, de coordenadas N 119.583,44m e E 410.879,79m, daí, segue em linha seca com o azimute plano 115°49'48" e distância de 893,4m até o ponto P-24, de coordenadas N 119.194,17m e E 411.683,91m, daí, segue em linha seca com o azimute plano 189°8'24" e distância de 1.415,6m até o ponto P-25, de coordenadas N 117.796,60m e E 411.459,00m, daí, segue em linha seca com o azimute plano 232°4'12" e distância de 976,4m até o ponto P-26, de coordenadas N 117.196,36m e E 410.688,84m, daí, segue em linha seca com o azimute plano 280°13'48" e distância de 946,8m até o ponto P-27, de coordenadas N 117.364,55m e E 409.757,15m, daí, segue em linha seca com o azimute plano 330°45'0" e distância de 724,5m até o ponto P-28, de coordenadas N 117.996,66m e E 409.403,14m, daí, segue em linha seca com o azimute plano 354°6'36" e distância de 772,2m até o ponto P-29, de coordenadas N 118.764,82m e E 409.323,92m, daí, segue em linha seca com o azimute plano 62°15'0" e distância de 1.758,1m retornando-se ao ponto P-23 início desta descrição, fechando o polígono.

Todas as coordenadas aqui descritas estão georreferenciadas ao Sistema de Referência Geocêntrico para as Américas, SIRGAS 2000, e projetadas no sistema de projeção UTM (Universal Transversa de Mercator), com o Meridiano Central -51° WGr, fuso 22N. A área, o perímetro, azimutes, distâncias foram calculados no plano de projeção UTM.

Contrato de Concessão Florestal nº 01/2021 – Anexo 1 – Página 14 de 15

Contrato de Concessão Florestal nº 01/2021 – Anexo 1 – Página 15 de 15

## Capturando Padrões

Utilizaremos expressões regulares (regex) para extração das informações de interesse, a saber: Código de cada vértice, coordenadas geodésicas UTM referente ao leste e norte. Ao observamos o memorial descritivo acima percebemos que há um padrão definido para os códigos dos vértices, cada um inicia com a letra maiúscula `P` seguida de um hífen e de um número com dois dígitos, por exemplo o primeiro vértice é denominado como `P-01`, o segundo com `P-02` o terceiro como `P-03` e assim por diante até o vértice `P-29`. A expressão regular a seguir captura todos os vértices disponíveis no memorial descritivo:

```
In [11]: # importar o pacote "re" para usar com regex
import re

# Regex para extração dos nomes dos vértices
vert = re.findall("[A-Z]+\d{2}", pdf_txt)
print(vert)
```

['P-01', 'P-02', 'P-03', 'P-04', 'P-05', 'P-06', 'P-07', 'P-08', 'P-09', 'P-10', 'P-11', 'P-12', 'P-13', 'P-14', 'P-15', 'P-17', 'P-18', 'P-19', 'P-20', 'P-21', 'P-01', 'P-23', 'P-24', 'P-25', 'P-26', 'P-27', 'P-28', 'P-29', 'P-23']

Vemos acima que a expressão não conseguiu extrair o nome do vértice 16, mas sabemos que os códigos são uma sequência de 1 até 29, podendo serem gerados apartir do seguinte código:

```
In [47]: vert = []

for i in range(1, 10):
    vert.append('P-0'+str(i))

for i in range(10, 30):
    vert.append('P-'+str(i))
```

```
In [48]: print(vert)
print()
print(len(vert))
```

['P-01', 'P-02', 'P-03', 'P-04', 'P-05', 'P-06', 'P-07', 'P-08', 'P-09', 'P-10', 'P-11', 'P-12', 'P-13', 'P-14', 'P-15', 'P-16', 'P-17', 'P-18', 'P-19', 'P-20', 'P-21', 'P-22', 'P-23', 'P-24', 'P-25', 'P-26', 'P-27', 'P-28', 'P-29']

29

Agora resta capturar o padrão das coordenadas UTM Leste e Norte. Pelo fato da área do memorial descritivo está localizada no hemisfério norte, mais precisamente no município de Ferreira Gomes, esatado do Amapá, ambas coordenadas apresentam-se com o mesmo número de dígitos. Em termos de padrão textual a coordenada norte sempre inicia com a letra maiúscula "N", 3 dígitos seguidos de um ponto, mais 3 dígitos, uma vígula e por fim dois dígitos e a letra minúscula "m", a exemplo da coordenada norte referente ao vértice P-29, N 118.764,82m. A coordenada leste possui o mesmo padrão com a diferença que sempre inicia com a letra maiúscula E, E 409.323,92m. De posse de tal padrão podemos aplicar uma expressão regular para extrair estas coordenadas.

```
In [42]: norte = re.findall(r'N\s*(\d{3}\.\d{3},\d{2})m\b', pdf_txt)
print(norte)
print(len(norte))
```

['140.443,12', '139.245,91', '137.260,28', '138.132,31', '130.563,09', '131.013,35', '130.950,90', '130.796,93', '126.064,02', '125.490,99', '123.506,24', '124.133,18', '122.731,73', '122.740,51', '122.286,80', '114.544,32', '111.841,13', '109.808,42', '115.818,01', '117.729,48', '117.729,80', '120.389,80', '120.672,98', '119.583,44', '119.194,17', '117.796,60', '117.196,36', '117.364,55', '117.996,66', '118.764,82']

29

```
In [43]: leste = re.findall(r'E\s*(\d{3}\.\d{3},\d{2})m\b', pdf_txt)
print(leste)
print(len(leste))
```

['410.769,20', '415.407,42', '416.613,41', '418.765,81', '423.378,13', '423.560,87', '428.236,94', '428.294,17', '426.838,05', '424.926,37', '423.091,73', '422.056,57', '419.643,04', '417.505,46', '416.776,58', '411.181,91', '412.251,32', '412.385,57', '399.342,22', '400.652,90', '400.773,23', '400.212,53', '410.879,79', '411.683,91', '411.459,00', '410.688,84', '409.757,15', '409.403,14', '409.323,92']

29

## Gerar um dataframe com os dados e salvar como .csv

Agora que temos lista dos vértices, coordenadas leste e norte, podemos juntá-las em um *dataframe* e salvá-lo no formato `.csv` ou `.xls`.

```
In [49]: # Importar a biblioteca pandas
import pandas as pd

# gerar o dataframe
df = pd.DataFrame({
    'vertice': vert,
    'leste': leste,
    'norte': norte
})
```

```
In [57]: # Salvar o data frame no disco nos formatos .csv e .xlsx
df.to_csv('./vertices_test.csv', index=False, sep=',')
df.to_excel('./vertices_test.xlsx', index=False)
```

In [ ]: