```
import requests
from bs4 import BeautifulSoup
import pandas as pd
import re

html = """<html lang="es">
<head>
   <meta charset="UTF-8">
   <title>Página de prueba</title>
</head>
<body>
<div id="main" class="full-width">
   <h1>El título de la página</h1>
   <p>Este es el primer párrafo</p>
   <p>Este es el segundo párrafo</p>
   <div id="innerDiv">
      <div class="links">
         <a href="https://pagina1.xyz/">Enlace 1</a>
         <a href="https://pagina2.xyz/">Enlace 2</a>
      </div>
      <div class="right">
         <div class="links">
            <a href="https://pagina3.xyz/">Enlace 3</a>
            <a href="https://pagina4.xyz/">Enlace 4</a>
         </div>
      </div>
   </div>
   <div id="footer">
      <!-- El footer -->
      <p>Este párrafo está en el footer</p>
      <div class="links footer-links">
         <a href="https://pagina5.xyz/">Enlace 5</a>
      </div>
   </div>
</div>
</body>
</html>"""
soup=BeautifulSoup(html,"html.parser")   Esto es necesario para que se lea el código

title=soup.title.string
title
'Página de prueba'
title_lista=[title]
title_lista
['Página de prueba']
title_lista1=title_lista*6
title_lista1
['Página de prueba',
 'Página de prueba',
 'Página de prueba',
 'Página de prueba',
 'Página de prueba',
 'Página de prueba']
id1=[ ]
for tag in soup.find_all(True,{"id":True}):  #Para tag que se id en soup encuentra todos los que hay en el texto html.
   id1.append(tag["id"])
id1
['main', 'innerDiv', 'footer']
id1.pop(1)
id1
['main', 'footer']
id1.insert(1,"Nan")
id1
['main', 'Nan', 'footer']
id1.insert(2,"Nan")
id1

['main', 'Nan', 'Nan', 'footer']
id1.insert(3,"Nan")
id1
['main', 'Nan', 'Nan', 'Nan', 'footer']
id1.insert(4,"footer")
id1
```

```
['main', 'Nan', 'Nan', 'Nan', 'footer', 'footer']
ID=id1
ID
['main', 'Nan', 'Nan', 'Nan', 'footer', 'footer']
dir=soup.find_all("a")
dir

[<a href="https://pagina1.xyz/">Enlace 1</a>,
 <a href="https://pagina2.xyz/">Enlace 2</a>,
 <a href="https://pagina3.xyz/">Enlace 3</a>,
 <a href="https://pagina4.xyz/">Enlace 4</a>,
 <a href="https://pagina5.xyz/">Enlace 5</a>]
Lst_url=[]
for url in dir:
    link = url.get('href')
    Lst_url.append(link)
Lst_url
['https://pagina1.xyz/',
 'https://pagina2.xyz/',
 'https://pagina3.xyz/',
 'https://pagina4.xyz/',
 'https://pagina5.xyz/']
pos=["p","a","a","a","div","p"]
length=len(pos)
length
tag=[]
for i in range(length):
    tag_1=(pos[i])
    tag.append(tag_1)
print(tag)
['p', 'a', 'a', 'a', 'div', 'p']
parraf=soup.find_all("p")
parraf2=parraf[1::]
parraf2
[<p>Este es el segundo párrafo</p>, <p>Este párrafo está en el footer</p>]
div=soup.find_all(class_="links footer-links")
div=str(div)
div2=div[13:31]
div2
'links footer-links'
values=Lst_url+parraf2
values
['https://pagina1.xyz/',
 'https://pagina2.xyz/',
 'https://pagina3.xyz/',
 'https://pagina4.xyz/',
 'https://pagina5.xyz/',
 <p>Este es el segundo párrafo</p>,
 <p>Este párrafo está en el footer</p>]
values1=(values[5])
values1
<p>Este es el segundo párrafo</p>
values2=values[0]
values2
'https://pagina1.xyz/'
values3=values[3]
values3
'https://pagina4.xyz/'
values4=values[4]
values4
'https://pagina5.xyz/'
values5=(values[6])
values5
<p>Este párrafo está en el footer</p>
type(values5)
bs4.element.Tag
total_values=[values[5]]+[values[0]]+[values[3]]+[values[4]]+[div2]+[values5]

total_values

[<p>Este es el segundo párrafo</p>,

 'https://pagina1.xyz/',
```

'https://pagina4.xyz/',

'https://pagina5.xyz/',

'links footer-links',

Este párrafo está en el footer]

print(title_lista1)

print(ID)

print(tag)

print(total_values)

['Página de prueba', 'Página de prueba', 'Página de prueba', 'Página de prueba', 'Página de prueba', 'Página de prueba']

['main', 'Nan', 'Nan', 'Nan', 'footer', 'footer']

['p', 'a', 'a', 'a', 'div', 'p']

[Este es el segundo párrafo , 'https://pagina1.xyz/', 'https://pagina4.xyz/', 'https://pagina5.xyz/', 'links footer-links', Este párrafo está en el footer]

data={"col1": title_lista1,"col2":ID,"col3":tag,"col4":total_values,

}

df=pd.DataFrame(data)

df

	col1	col2	col3	col4
0	Página de prueba	main	p	[Este es el segundo párrafo]
1	Página de prueba	Nan	a	https://pagina1.xyz/
2	Página de prueba	Nan	a	https://pagina4.xyz/
3	Página de prueba	Nan	a	https://pagina5.xyz/
4	Página de prueba	footer	div	links footer-links
5	Página de prueba	footer	p	[Este párrafo está en el footer]

df.columns = ["Title","ID","Tags","Values"]

df

	Title	ID	Tags	Values
0	Página de prueba	main	p	[Este es el segundo párrafo]
1	Página de prueba	Nan	a	https://pagina1.xyz/
2	Página de prueba	Nan	a	https://pagina4.xyz/
3	Página de prueba	Nan	a	https://pagina5.xyz/
4	Página de prueba	footer	div	links footer-links
5	Página de prueba	footer	p	[Este párrafo está en el footer]