

- Estudo de padrão de Xpath resultados do google search
  - primeiro os primeiros 11 resultados da pesquisa
  - Após os primeiros 10 resultados começamos a coletar por barra de rolagem, vamos mapeando novas divs
  - Autor

# Estudo de padrão de Xpath resultados do google search

---

## primeiro os primeiros 11 resultados da pesquisa

Eles aparecem em uma div com o ID "search" e cada resultado em particular estará dentro de 3 novas divs aninhadas passando de 1 a 11 para pegar as primeiras 11 respostas que não contenham anúncios.

```
# Div mãe:
//*[@id="search"]

# Div de cada resultado da pesquisa:
//*[@id="search"]/div/div/div[1]
//*[@id="search"]/div/div/div[11]
```

## Após os primeiros 10 resultados começamos a coletar por barra de rolagem, vamos mapeando novas divs

com base no seguinte os resultados adicionais aparecem em divs com o ID nomeados no padrão "**arc-srp\_#**" em que # pode ser um número começando por 110 e a cada nova rolagem entram mais 10 na contagem até chegar ao 190 depois disso vai para 1100 e depois 1110, 1120, 1130...1190, 1200,1210...1990,10100. consegue entender o padrão ? Com base nisso podemos desenhar nosso fluxo de coleta em uma consulta infinitamente, até o limite de resultados, aliás o limite de resultados é o valor aproximado de retorno e a capacidade da sua memória da máquina, o que chegar ao

limite primeiro. e o número aproximado de retornos encontramos no caminho xpath :

`//*[@id="result-stats"]` é esse mesmo :D

então desenhando um padrão de coleta pelos resultados adicionais. primeiro conjunto de resultados adicionais, por algum motivo os resultados começam a partir da 3 div da rolagem e vão até

```
# div mãe:
//*[@id="arc-srp_110"]
# Div de cada resultado da pesquisa:
//*[@id="arc-srp_110"]/div/div[3]/div/div < - primeira
//*[@id="arc-srp_110"]/div/div[13]/div/div < - última
# div mãe:
//*[@id="arc-srp_120"]
# Div de cada resultado da pesquisa:
//*[@id="arc-srp_120"]/div/div[3]/div/div < - primeira
//*[@id="arc-srp_120"]/div/div[12]/div/div < - última
..
# div mãe:
//*[@id="arc-srp_1100"]
# Div de cada resultado da pesquisa:
//*[@id="arc-srp_1100"]/div/div[3]/div/div < - primeira
//*[@id="arc-srp_1100"]/div/div[12]/div/div < - última
```

Obs: note uma outra variante, após o segundo conjunto de resultados da barra de rolagem, passam a vir 9 resultados sem ser anúncio, neste contexto vai de 3 à 12 totalizando 9 retornos por rolagem.

## Autor

- Gustavo H Martins ([GitHub](#) | [LinkedIn](#))



Essas explicações ajudam a compreender como os scripts funcionam e como podem ser integrados em outros projetos.