AAG07

1. Abrir o arquivo em anexo
2. Ler as colunas do scholar e do scopus
3. Para cada professor, pegar o número total de citações e o h-index
4. Gerar uma saida com "Nome - total scholar - hindex scholar - total scopus - hindex scopus

ps: o arquivo de saída pode ser separado por vírgula

OBS.: Fiz uma correção no arquivo original pois o link do Scopus do Alex estava incorreto, estava igual ao do André.

In [62]:
```
import pandas as pd

# Carregar o arquivo
file_path = 'corpo Docente - corrigido.xlsx'
data = pd.ExcelFile(file_path)
data.sheet_names
```

Out[62]:
```
['Sheet1']
```

In [63]:
```
# Carregar Sheet1
df = data.parse('Sheet1')
df.head(1)
```

Out[63]:

	Unnamed: 0	ORCID	SCHOLAR	
0	Alex Borges Vieira	https://orcid.org/0000-0003-0821-126X	https://scholar.google.com.br/citations?user=t...	https://www.scopus.com/authid/

In [64]:
```
# Renomear a coluna do Nome
df.rename(columns={'Unnamed: 0': 'NOME'}, inplace=True)
df.head(1)
```

Out[64]:

	NOME	ORCID	SCHOLAR	SC
0	Alex Borges Vieira	https://orcid.org/0000-0003-0821-126X	https://scholar.google.com.br/citations?user=t...	https://www.scopus.com/authid/det a

Extrair os dados do Scholar

In [65]:
```
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.service import Service
from selenium.webdriver.chrome.options import Options
from webdriver_manager.chrome import ChromeDriverManager
import time

# Carregue os dados
links = df['SCHOLAR']

# Configuração do Selenium
```

```
chrome_options = Options()
chrome_options.add_argument("--headless")
driver = webdriver.Chrome(service=Service(ChromeDriverManager().install()), \
    options=chrome_options)

# Listas para armazenar os resultados
citations = []
h_indexes = []

for link in links:
    try:
        # Acesse o link do Google Scholar
        driver.get(link)
        time.sleep(2)  # Aguarde o carregamento da página

        print(link)

        # Extrir o número de citações
        citation_element = driver.find_element(By.XPATH, '(//td[@class="gsc_rsb_std"])[1]')
        citations.append(citation_element.text)

        # Extrir o h-index
        h_index_element = driver.find_element(By.XPATH, '(//td[@class="gsc_rsb_std"])[3]')
        h_indexes.append(h_index_element.text)

        print(citation_element.text, h_index_element.text)

    except Exception as e:
        citations.append("Erro")
        h_indexes.append("Erro")
        print(f"Erro ao processar o link: {link}\n{e}")

# Feche o navegador
driver.quit()

# Salve os resultados no DataFrame
df['SCHOLAR CITATIONS'] = citations
df['SCHOLAR H-INDEX'] = h_indexes

df.head()
```

```
https://scholar.google.com.br/citations?user=tjvh6lkAAAAJ&hl=en
2536 23
https://scholar.google.com.br/citations?user=S4cwYUwAAAAJ&hl=en&oi=ao
442 10
https://scholar.google.com.br/citations?user=V_ePJHYAAAAJ&hl=en&oi=ao
488 11
https://scholar.google.com.br/citations?user=8fURLscAAAAJ&hl=en
466 12
https://scholar.google.com.br/citations?user=YufvCREAAAAJ&hl=en&oi=ao
1813 18
https://scholar.google.com.br/citations?user=7_-SsxwAAAAJ&hl=en&oi=ao
1436 19
https://scholar.google.com.br/citations?user=CiZNGSQAAAAJ&hl=en&oi=ao
760 12
https://scholar.google.com.br/citations?hl=en&user=nP1a1zgAAAAJ
1311 16
https://scholar.google.com.br/citations?user=hxW0lxMAAAAJ&hl=en&oi=ao
106 5
https://scholar.google.com.br/citations?hl=en&user=LpGANt0AAAAJ
694 13
https://scholar.google.com.br/citations?hl=en&user=Tr8Bp-AAAAAJ
399 9
https://scholar.google.com.br/citations?user=YDgkwJwAAAAJ&hl=en&oi=ao
2186 22
https://scholar.google.com.br/citations?user=bP7R6tQAAAAJ&hl=en&oi=ao
776 10
https://scholar.google.com.br/citations?user=j4UDSXgAAAAJ&hl=en&oi=ao
197 8
https://scholar.google.com.br/citations?user=tgyrRMkAAAAJ&hl=en
1880 21
https://scholar.google.com.br/citations?user=PJpvvooAAAAJ&hl=en&oi=ao
244 8
https://scholar.google.com.br/citations?user=UF79ycEAAAAJ&hl=en&oi=ao
142 7
https://scholar.google.com.br/citations?user=n3A_2JYAAAAJ&hl=en&oi=sra
906 14
https://scholar.google.com.br/citations?user=Of35aZYAAAAJ&hl=en&oi=ao
299 11
```

Out[65]:

	NOME	ORCID	SCHOLAR	
0	Alex Borges Vieira	https://orcid.org/0000-0003-0821-126X	https://scholar.google.com.br/citations?user=t...	https://www.scopus.com/authic
1	André Luiz de Oliveira	https://orcid.org/0000-0003-0564-0034	https://scholar.google.com.br/citations?user=S...	https://www.scopus.com/authic
2	Carlos Cristiano Hasenclever Borges	https://orcid.org/0000-0001-7413-2880	https://scholar.google.com.br/citations?user=V...	https://www.scopus.com/authic
3	Edelberto Silva	https://orcid.org/0000-0002-0058-9260	https://scholar.google.com.br/citations?user=8...	https://www.scopus.com/authic
4	Fernanda Cláudia Alves Campos	https://orcid.org/0000-0002-0763-2698	https://scholar.google.com.br/citations?user=Y...	https://www.scopus.com/authic

Extrair os dados do Scopus

Tive que retirar a virgula dos números para evitar erros.

In [66]:

```
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.service import Service
from selenium.webdriver.chrome.options import Options
from webdriver_manager.chrome import ChromeDriverManager
import time

# Configuração do Selenium
chrome_options = Options()
chrome_options.add_argument("--start-maximized")
driver = webdriver.Chrome(service=Service(ChromeDriverManager().install()), \
    options=chrome_options)

# Carregue os dados
links = df['SCOPUS']

# Listas para armazenar os resultados
citations = []
h_indexes = []

for link in links:
    try:
        # Acesse o link do Google Scholar
        driver.get(link)
        time.sleep(2)  # Aguarde o carregamento da página

        print(link)

        # Extrir o número de citações
        citation_element = driver.find_element(By.XPATH, \
            '(//span[@data-testid="unclickable-count"])[1]')
        citations.append(citation_element.text.replace(",", ""))

        # Extrir o h-index
        h_index_element = driver.find_element(By.XPATH, \
            '(//span[@data-testid="unclickable-count"])[3]')
        h_indexes.append(h_index_element.text.replace(",", ""))

        print(citation_element.text, h_index_element.text)

    except Exception as e:
        citations.append("Erro")
        h_indexes.append("Erro")
        print(f"Erro ao processar o link: {link}\n{e}")

# Feche o navegador
driver.quit()

# Salve os resultados no DataFrame
df['SCOPUS CITATIONS'] = citations
df['SCOPUS H-INDEX'] = h_indexes

df.head()
```

```
https://www.scopus.com/authid/detail.uri?authorId=35079585800
1,429 18
https://www.scopus.com/authid/detail.uri?authorId=57197404486
219 8
https://www.scopus.com/authid/detail.uri?authorId=7005937032
241 9
https://www.scopus.com/authid/detail.uri?authorId=24315363000
207 8
https://www.scopus.com/authid/detail.uri?authorId=35319797200
441 11
https://www.scopus.com/authid/detail.uri?authorId=23974282400
819 14
https://www.scopus.com/authid/detail.uri?authorId=57212499803
245 8
https://www.scopus.com/authid/detail.uri?authorId=18433696900
502 11
https://www.scopus.com/authid/detail.uri?authorId=59158282700
61 4
https://www.scopus.com/authid/detail.uri?authorId=7202140734
334 10
https://www.scopus.com/authid/detail.uri?authorId=15061649700
170 6
https://www.scopus.com/authid/detail.uri?authorId=9744594600
927 14
https://www.scopus.com/authid/detail.uri?authorId=56878649700
290 7
https://www.scopus.com/authid/detail.uri?authorId=23976522600
104 8
https://www.scopus.com/authid/detail.uri?authorId=7006111233
622 12
https://www.scopus.com/authid/detail.uri?authorId=36183040300
137 7
https://www.scopus.com/authid/detail.uri?authorId=57204592348
49 4
https://www.scopus.com/authid/detail.uri?authorId=35191111500
377 10
https://www.scopus.com/authid/detail.uri?authorId=6508089106
125 8
```

Out[66]:

	NOME	ORCID	SCHOLAR	
0	Alex Borges Vieira	https://orcid.org/0000-0003-0821-126X	https://scholar.google.com.br/citations?user=t...	https://www.scopus.com/authic
1	André Luiz de Oliveira	https://orcid.org/0000-0003-0564-0034	https://scholar.google.com.br/citations?user=S...	https://www.scopus.com/authic
2	Carlos Cristiano Hasenclever Borges	https://orcid.org/0000-0001-7413-2880	https://scholar.google.com.br/citations?user=V...	https://www.scopus.com/authic
3	Edelberto Silva	https://orcid.org/0000-0002-0058-9260	https://scholar.google.com.br/citations?user=8...	https://www.scopus.com/authic
4	Fernanda Cláudia Alves Campos	https://orcid.org/0000-0002-0763-2698	https://scholar.google.com.br/citations?user=Y...	https://www.scopus.com/authic

Gerar arquivo .csv com o resultado

In [67]:
```
df_final = df[['NOME', 'SCHOLAR CITATIONS', 'SCHOLAR H-INDEX', 'SCOPUS CITATIONS', 'SCOPUS H-
df_final.to_csv('corpo Docente - final.csv', index=False)
```