Hello World!

This script pulls the top headlines from the Drudge Report and saves them into a Pandas data frame. It also saves the time the script pulled them since Drudge changes a lot.

Best,
George John Jordan Thomas Aquinas Hayward, Optimist

In [2]:
```python
from bs4 import BeautifulSoup
import requests
import pandas as pd
import datetime
from datetime import date

webpage_response = requests.get('https://www.drudgereport.com/')
webpage = webpage_response.content
soup = BeautifulSoup(webpage, "html.parser")
#the key here is that Drudge's top headlines are under the 'tt' tag
headers_raw = soup.tt.get_text()
#this is a sensible place to split
split_up = headers_raw.split("\n")

#here, we just remove the spaces that get left in the split-up list
lines = []
for i in split_up:
    if i != '':
        lines.append(i)

#we make a dictionary to help us sort through what we have and prepare for moving to a pandas dataframe
drudge_headlines = {}
drudge_headlines['top_headline'] = lines
#drudge headlines change all the time, so we'll want to know the day and the datetime of this scrape
drudge_headlines['date'] = date.today()
drudge_headlines['time'] = datetime.datetime.now()

#pandas
Drudge_Report = pd.DataFrame.from_dict(drudge_headlines)

#getting the date into the file_name
datetime_now_string = datetime.datetime.now().strftime("%m-%d-%Y_%H-%M-%p")
Drudge_Report.to_csv("drudge_report_"+datetime_now_string+".csv",index=False)
#!!--->I prefer to write to Excel for my own purposes, and if you don't have openpyxl, pandas won't write it
#"pip install openpyxl"
Drudge_Report.to_excel("drudge_report_"+datetime_now_string+".xlsx",index=False)

#here's the dataframe
Drudge_Report
```

Out[2]:

	top_headline	date	time
0	G20: Beneath smiles and handshakes, tensions s...	2019-06-29	2019-06-29 01:08:28.719903
1	Saudi crown prince feted at summit...	2019-06-29	2019-06-29 01:08:28.719903
2	FRANCE HOTTEST DAY IN HISTORY!	2019-06-29	2019-06-29 01:08:28.719903