import regex as re

import time

import pandas as pd

import requests

from bs4 import BeautifulSoup

from selenium import webdriver

from selenium.webdriver.chrome.service import Service

from selenium.webdriver.common.by import By

from selenium.webdriver.common.keys import Keys as KeysBrowser

from webdriver\_manager.chrome import ChromeDriverManager

from matplotlib import pyplot as plt

**# Initialize the Chrome driver**

driver = webdriver.Chrome(service=Service(ChromeDriverManager().install()))

# Load the search page

driver.get("https://www.bailii.org/form/search\_cases.html")

# Find the search box element by name and enter the search query for "Companies Act 2006"

search\_box = driver.find\_element(By.NAME, "phrase")

search\_box.send\_keys("Companies Act 2006")

search\_box.send\_keys(KeysBrowser.RETURN)

**# Define search terms and patterns**

search\_terms = ["Section 172", "Sections 172", "section 172", "sections 172", "ss 172", "SS 172"]

keywords\_pattern = re.compile("|".join(search\_terms))

context\_pattern = re.compile(r"(?s)(.{0,200})(Companies Act 2006)(.{0,200})")

**# Collect the URLs for all the search results**

citation\_urls = []

while True:

    soup = BeautifulSoup(driver.page\_source, 'html.parser')

    ol = soup.find\_all("ol")[0]  # Collect the ol element from the page

    if not ol:

        print("No ol elements found")

        break

    for li in ol.find\_all("li"):

        citation\_link = li.find("a")

        if citation\_link:

            citation\_url = "https://www.bailii.org/" + citation\_link["href"]

            citation\_urls.append(citation\_url)

    try:

        # Check if the "Next 10" button exists

        next\_button = driver.find\_element(By.XPATH, "//input[@value='Next 10 >>>']")

    except:

        # If the button is not found, break out of the loop

        break

    next\_button.click()  # Click the "Next 10" button

    time.sleep(2)  # Wait for the next page to load

**# Collect the URLs for the relevant search results**

relevant\_urls = []

for url in citation\_urls:

    # Load the page for each URL

    driver.get(url)

    soup = BeautifulSoup(driver.page\_source, 'html.parser')

    text = soup.get\_text()

    # Search for the relevant keywords and context within a specific range of characters

    matches = context\_pattern.finditer(text)

    for match in matches:

        if match and (keywords\_pattern.search(match.group(1)) or keywords\_pattern.search(match.group(3))):

            relevant\_urls.append(url)

            break

**# Print the relevant URLs and the total number of relevant URLs**

print(relevant\_urls)

print(len(relevant\_urls))

**Backend MacOSX is interactive backend. Turning interactive mode on.**

[WDM] - Downloading: 100%|██████████| 8.84M/8.84M [01:05<00:00, 141kB/s]

['https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2008/2198.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2008/1534.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2022/855.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2022/1110.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2020/2839.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2010/577.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2015/2540.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2019/112.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2019/2304.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2020/1932.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/Misc/2021/7.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2019/1358.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2020/2536.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/Misc/2021/7.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/nie/cases/NIHC/Ch/2020/8.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2017/2621.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2018/1793.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Comm/2013/3615.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2021/22.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2020/2808.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2021/2661.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2013/2631.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2018/2982.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/nie/cases/NIHC/Ch/2008/23.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2019/3338.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/nie/cases/NIHC/Ch/2008/23.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/QB/2008/2340.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/QB/2008/2340.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2018/2807.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2019/90.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2020/54.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/QB/2021/2103.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/QB/2021/2103.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2018/756.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/uk/cases/UKSC/2021/49.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/QB/2018/1203.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/QB/2018/1203.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2020/3533.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/uk/cases/UKFTT/TC/2019/TC07033.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/scot/cases/ScotCS/2013/2013CSOH181.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2017/126.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/scot/cases/ScotCS/2013/2013CSOH181.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2022/1233.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2018/3201.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2019/1664.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2022/2324.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/uk/cases/UKSC/2015/23.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2022/1349.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2021/787.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2012/61.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2012/61.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2015/2258.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2017/2767.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2017/3228.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/uk/cases/CAT/2017/8.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/QB/2018/1506.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2019/459.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2022/1028.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/uk/cases/UKFTT/TC/2020/TC07881.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2022/364.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/uk/cases/UKFTT/TC/2020/TC07881.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/scot/cases/ScotCS/2021/2021\_CSOH\_109.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2019/111.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2022/291.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2011/1168.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2022/1099.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2011/1168.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2020/1626.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2018/472.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2021/1902.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2019/485.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/TCC/2022/884.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/QB/2019/843.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/QB/2019/843.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2020/2649.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2020/2913.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2022/153.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2020/1674.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2020/2913.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/scot/cases/ScotCS/2018/[2018]\_CSOH\_42.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/uk/cases/CAT/2017/6.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2021/175.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2021/2600.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2018/84.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/TCC/2014/2725.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2018/1924.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2022/784.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2021/2550.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2022/3562.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2010/2029.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/uk/cases/UKFTT/TC/2022/TC08541.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2021/325.html&query=("Companies+Act+2006")', 'https://www.bailii.org//cgi-bin/format.cgi?doc=/ew/cases/EWHC/Comm/2022/935.html&query=("Companies+Act+2006")']

93

**# Collect relevant information from each URL**

dates = []

jurisdictions = []

judges = []

citations = []

for url in relevant\_urls:

    driver.get(url)

    soup = BeautifulSoup(driver.page\_source, 'html.parser')

    text = soup.get\_text()

    # Extract the date from the page

    date = soup.find("span", {"class": "docdate"})

    if date:

        dates.append(date.text.strip())

    else:

        dates.append(None)

    # Extract the jurisdiction from the page

    jurisdiction = soup.find("span", {"class": "jurisdiction"})

    if jurisdiction:

        jurisdictions.append(jurisdiction.text.strip())

    else:

        jurisdictions.append(None)

    # Extract the judge(s) from the page

    caseheader = soup.find("div", {"class": "caseheader"})

    if caseheader:

        judge = caseheader.find("strong")

        if judge:

            judges.append(judge.text.strip())

        else:

            judges.append(None)

    else:

        judges.append(None)

    # Extract the citations from the page

    if caseheader is not None:

        citation = caseheader.find\_all("span", {"class": "citation"})

        if citation:

            citations.append(', '.join([c.text.strip() for c in citation]))

        else:

            citations.append(None)

    else:

        citations.append(None)

# Create Pandas DataFrame

df = pd.DataFrame({

    "Date": dates,

    "Jurisdiction": jurisdictions,

    "Judge": judges,

    "Citation": citations

})

# Display DataFrame

print(df)

      Date Jurisdiction Judge Citation

0     None         None  None     None

1     None         None  None     None

2     None         None  None     None

3     None         None  None     None

4     None         None  None     None

    ...          ...   ...      ...

1677  None         None  None     None

1678  None         None  None     None

1679  None         None  None     None

1680  None         None  None     None

1681  None         None  None     None

[1682 rows x 4 columns]