

# WebCrawler Project Report

杨辉宗 - 1820232064

## Experiment Environment

- Operating System: MacOS Sequoia 15.1
- IDE: PyCharm CE
- Python version 3.11

## Installed Library

- Requests
- BeautifulSoup4

## Code Analysis

Import Lib

```
1 import requests
2 from bs4 import BeautifulSoup
```

Web crawling functions

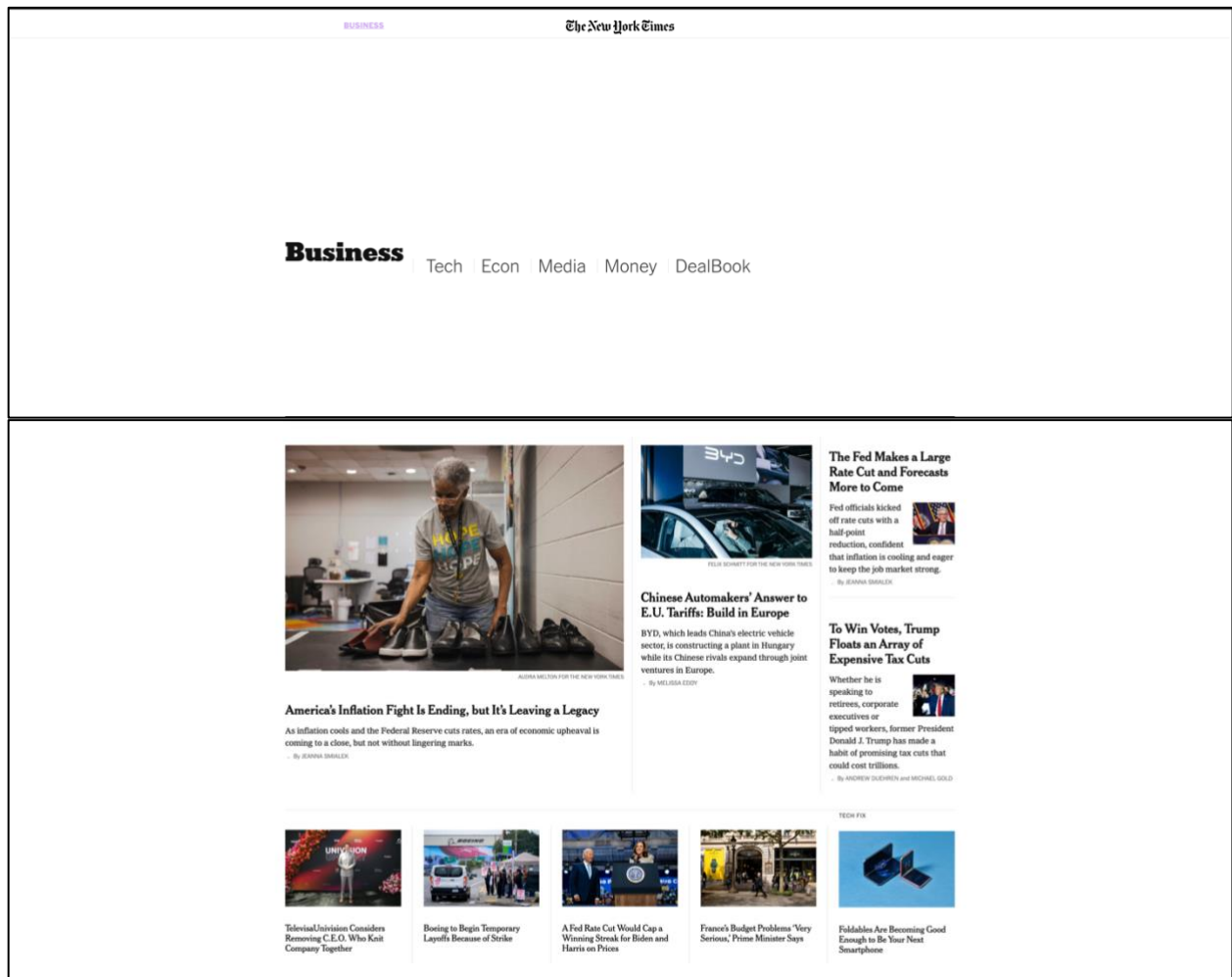
```
4 def crawl_web_page(url): 1 usage
5     headers = {
6         'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 '
7         '(KHTML, like Gecko) Chrome/58.0.3029.110 Safari/537.36'
8     }
9     try:
10         response = requests.get(url, headers=headers)
11         if response.status_code == 200:
12             soup = BeautifulSoup(response.content, features='html.parser')
13             data = soup.prettify() # This formats the HTML content nicely
14
15             # Print the data to console
16             print(data)
17
18             # Save the data to a file
19             file_path = 'output.html' # Define the local file path to store the data
20             with open(file_path, 'w', encoding='utf-8') as file:
21                 file.write(data)
22             print(f"Data saved to {file_path}")
23
24         else:
25             print(f"Failed to retrieve the page. Status code: {response.status_code}")
26     except Exception as e:
27         print(f"An error occurred: {e}")
```

Specify link to crawl and call the function to start

```
29 # Specify the webpage to crawl
30 url = 'https://www.nytimes.com/section/business'
31 crawl_web_page(url)
```

## Results

 output.html



## Latest

Search

### Amazon Joins Hollywood's Top Lobbying Group

The Motion Picture Association has been wooing the company for years. The addition gives the group a lot more muscle.

By BROOKS GARNES



#### DEALBOOK NEWSLETTER

### What the Fed's Big Rate Cut Reveals About the Economy

Investors have sent stocks and bonds higher on expectations of a soft landing. But some Republicans think the central bank overstepped its bounds.

By ANDREW ROSS SORIN, RAY MATTL, BERNARD WANNER, SARAH KESSLER, MICHAEL J. DE LA MERCEZ, LAUREN HIRSCH and EPHRAIM LIND



### Bank of England Holds Rates Steady but Signals Gradual Easing

The British central bank has emphasized its intention to move steadily to ensure that inflation continues to slow.

By ERIK NELSON



### Axel Springer Strikes Deal With KKR to Split Up Publishing Giant

The deal values all of Axel Springer at roughly \$15 billion, with the company's publishing assets worth about \$4 billion.

By BENJAMIN MULLIN and LAUREN HIRSCH



### After Fed Cuts Rates, Biden Will Claim Credit for Economy's Strength

The president's speech on Thursday won't be a "victory lap," officials said, but it will celebrate falling inflation and borrowing costs along with solid growth.

By JIM TANKERSLEY



#### TRIPPED UP

### Help! My \$1,988 Spring Break Rental Was Canceled at the Last Minute.

A family reserved an eight-night Vibe in the U.S. Virgin Islands, but the owner canceled the day before the trip. Finding a comparable place was stressful and cost over \$4,000 more; Vibe wouldn't make up the full difference.

By SETH RUGEL



### Company Says It Is Investigating Radios Targeted in Lebanon Blasts

The Japanese electronics manufacturer Icom said it stopped making the walkie-talkie model in 2014 and has warned about fake versions for several years.

By RIVER ANNA DAVIS



### Valarie D'Elia, Travel Reporter on TV and Radio, Dies at 64

She steered vacationers and business travelers to choice destinations, talked about the best deals, and offered up savvy tips on how to avoid vacation.

By SAM ROBERTS



### Musk Finds a (Temporary) Way Around Brazil's X Ban

Elon Musk's social network used a technical maneuver to restore service for many Brazilians after a court blocked it. Regulators expected to restore the block soon.

By JACK NICAS

[Leer en español](#)



#### TIMES VIDEO

### Federal Reserve Cuts Interest Rates for the First Time in Four Years

Jerome H. Powell, the Fed chair, said that the central bank would take future interest rate cuts "meeting by meeting" after lowering rates by a half percentage point, an unusually large move.

By THE NEW YORK TIMES



#### BUSINESS DELIVERED TO YOUR INBOX

Sign up for one of our daily or weekly newsletters.

##### DealBook

The news driving the markets and the conversation. (Daily)

SIGN UP

##### Your Money

Tips on retirement, paying for college, credit cards and the right way to invest. (Weekly)

SIGN UP