The Impact Analysis of COVID-19 on China Various Industries Using Crawler Technology and Data Visualization Technology

Charles Chen
School of Computer Science and Engineering
Minnan Normal University
Zhangzhou City, China
Charles.chen@mnnu.edu.cn
2440216302@qq.com

Ling Chen
School of Computer Science and Engineering
Minnan Normal University
Zhangzhou City, China
2536554160@qq.com

Mingjun Xiao
School of Computer Science and Engineering
Minnan Normal University
Zhangzhou City, China
1055289315@qq.com

Jinfeng Ning
School of Computer Science and Engineering
Minnan Normal University
Zhangzhou City, China
1874145775@qq.com

2020 IEEE 3rd International Conference of Safe Production and Informatization (IICSPI) | 978-1-7281-7738-0/20/$31.00 ©2020 IEEE | DOI: 10.1109/IICSPI51290.2020.9332455

***Abstract*—In order to stop the spread of COVID-19, government has to lockdown of the city from all outside contact. Most of business activities force to stop. This will have an impact on the economic activities. In this paper, first we will use the crawler technology to fetch company's data from Cninfo website. Second, we will use the big data analysis and data visualization technology to analyze the impact of Covid-19 on China various industries. In comparing various industries' profit in the first quarter of 2019 and 2020, we find that except for agriculture, forestry, animal husbandry and fishery, national defense and military, banking, food and beverage industries, the profit value of other industries has declined, with transportation, chemical industry, mining and non-banking financial sector falling the most. On the whole, various industries have been hit by the epidemic.**

Keywords—COVID-19, Crawler technology, Data analysis, Data visualization, Selenium framework

I. Introduction

Since December 2019, novel coronavirus infection cases of acute respiratory tract infection have been confirmed by a number of cases of unknown pneumonia in Wuhan, Hubei, which have been found in many cases with exposure history of Southern China seafood market [1-3]. This novel coronavirus has been named as Corona Virus Disease 2019 (COVID-19) by WHO on February 12, 2020 [4]. Until August 23, 2020,there are 23334458 confirmed cases and 807154 deaths in the world. Compared with the previous day, 265806 new confirmed cases and 5191 new deaths occurred in a single day [5].

In order to prevent the spread of coronavirus, many governments have to lockdown the city from all outside contact. Various business activities are forced to stop. This will have an impact on the economic growing. According to the data released by the National Bureau of statistics on March 16, from January to February 2020, the epidemic will affect various indicators of China's economy to varying degrees. Industrial production and service industry production will decline, market consumption and investment will decrease, and the purchasing manager's index (PMI) will decline [6].

In order to accurately understand the influence of COVID-19 on China's economy, first, we will use the crawler technology to crawl company's data (the first quarter of 2019 and first quarter of 2020) from Cninfo website (http://www.cninfo.com.cn/new/index). Second, we will use the data visualization technology to analyze the impact of Covid-19 on China various industries via comparing various industries' profit in the first quarter of 2019 and 2020.

The structure of this paper is scheduled as follows. In the second section, we will introduce the related works of big data, Crawler technology, and Data visualization. In the third section, we will introduce our research methods. The experiment result is in fourth section. The conclusion is stated in the final section.

II. Related Work

A. Crawler Technology

Web crawler is an information acquisition technology with the development of search engines and it is also an important part of the index engine grabbing system [7]. It is used to down load the webpages contents on the Internet to the local area. Essentially, the function of Internet crawler is to track the hyper link structure of the network, and combine its own capture strategy, thus achieving the collection of web resources information [8].

Jian and Qin [8] state that the architecture of web crawler has four modules. There are download, subject content extraction, webpage classifier, and link evaluation and selection, as shown in Figure1.

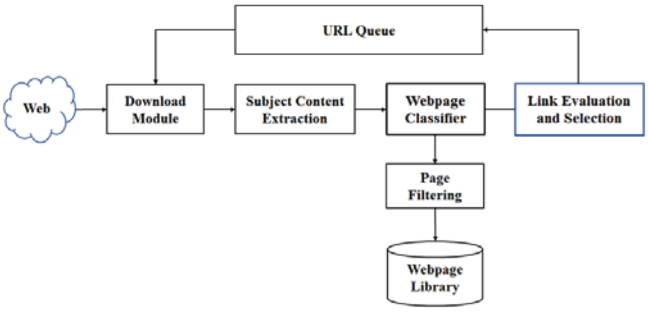


Fig. 1. The structure of network crawler.

Download module: The main work of the Web crawler is to download the web pages found. In the process of downloading, a link scheduling process is needed, which distributes links to all downloaded threads according to its own search strategy. When downloading webpages through web crawler system, each webpage has its own download threads [9].

Content extraction: The purpose of extracting web content is to select the links contained in the web pages. There are some methods to select the links such as displaying the web pages by structure tree or using the mode of pattern matching to select the links directly. In this process, it is necessary to delete the content that is irrelevant to the required information through the processing of “denoising” [9].

Webpage classifier: In this stage, the captured web pages will compare with the information needed to determine whether they meet the requirements. The user should provide information close to the subject as far as possible according to the general direction of the user [10].

Link selection and evaluation: In this final selection and evaluation stage, the advertising links affecting information content are removed first, and then the relative links are converted into absolute links to determine the final links. Finally, in the evaluation stage of links, those needed links will be placed in the queue to be crawled, and then unified evaluated.

B. Data Visualization Technology

Data visualization is a scientific and technological research on the visual representation of data. Among them, the visual representation of this kind of data is defined as a kind of information extracted in a certain summary form, including various attributes and variables of corresponding information units [11]. Data visualization is the combination of art and technology. It will be a variety of data graphical way to show people.

1) Basic concepts

Data visualization technology includes the following basic concepts [12].

- Data space: it is a multidimensional information space composed of n-dimensional attributes and m elements.
- Data development: it refers to the quantitative deduction and calculation of data by using certain algorithms and tools.
- Data analysis: it refers to the multi-dimensional data slice, block, rotation and other actions to analyze the data, so as to observe the data from multiple angles and sides.
- Data visualization: it refers to the process that the data in large data sets are represented in the form of graphics and images, and the unknown information is found by using data analysis and development tools.

2) Charts

Data visualization charts can be classified into the following categories according to the functions and functions of data: comparison, distribution, process, map, proportion, interval, association, time and trend. Each type of chart can contain different data visualization graphics, such as bar chart, pie chart, bubble chart, thermal chart, trend chart, histogram, radar chart, color block diagram, funnel chart, chord chart, dashboard, area chart, broken line chart, K-line chart, ring chart, word cloud, etc.

III. Research Methods

In our research methods, we will create a Python file using Selenium + BS4 framework to crawl the net profit of companies in Shenzhen main board, Shanghai main board and gem in the first quarter of 2019 and 2020. The URL address is: http://webapi.cninfo.com.cn/#/thematicStatistics. Then,we use Pandas, NumPy and Matplotlib libraries of Python to extract crawled data, calculate cleaned data and plot an visualizable graphic charts for analysis.

A. Crawler Framework and Python Library

1) Selenium

Selenium (https://www.selenium.dev/) is an open-source framework mainly used for testing web applications. It was originally an automated testing tool. But it is now used in crawlers to solve the problem that requests cannot directly execute JavaScript code. Selenium enables the impersonation of users interacting with browsers such as Chrome, Firefox, Edge, or Opera in an automated manner [13]. The essence of selenium is to drive the browser, fully simulate the browser's operation, such as jump, input, click, drop-down, etc., to get the result of web page rendering.

Selenium can be run on operating system platforms like Windows, Linux, and Macintosh etc. Selenium Suite is composed of following components: Selenium IDE, Selenium Core, Selenium 1 (also addressed as Selenium RC or Remote Control), Selenium 2 (also addressed as Selenium Web driver), and Selenium-Grid [14]. In our research, we will use the Selenium Web driver to crawler website.

2) Beautifulsoup(BS4)

Beautifulsoup is a HTML parsing tool of python library that can extract data from HTML or XML files. It provides some simple and python functions to handle navigation, search, and modify the analysis tree [15]. It can convert the input HTML or XML data into "Unicode" code, and the output HTML or XML data into "UTF-8" code. Therefore, it is not necessary to consider the data encoding mode when using the beautiful soup library to parse the data [16].

3) Matplotlib

Matplotlib is a Python 2D drawing library, which generates publishing quality graphics in various hard copy formats and cross platform interactive environment. With Matplotlib, developers can generate graphs, histograms, power spectra, bar graphs, error graphs, scatter charts, etc. with just a few lines of code [17].

4) NumPy

Numpy (numerical Python) is an open source numerical computation extension of Python. This tool can be used to store and process large matrices, which is much more efficient than Python's nested list structure (this structure can also be used to represent matrix), and supports a large number of dimensional arrays and matrix operations. In addition, it also provides a large number of mathematical function libraries for array operations [18].

5) Pandas

Pandas is a tool based on Numpy, which is created to solve data analysis tasks. Pandas includes a large number of libraries and some standard data models, providing the tools needed to operate large data sets efficiently. Functions and methods of Pandas can enable us to process data quickly and conveniently [19].

B. Crawler program coding and process flow

Before we start the coding, we have to use Python command to install Selenium (pip install Selenium) and BS4 (pip install beautifulsoup4) frame first. After system environment is built, we will use PyCharm (Python IDE) tool to code Python program. Figure 2 is the process flow of website crawler.

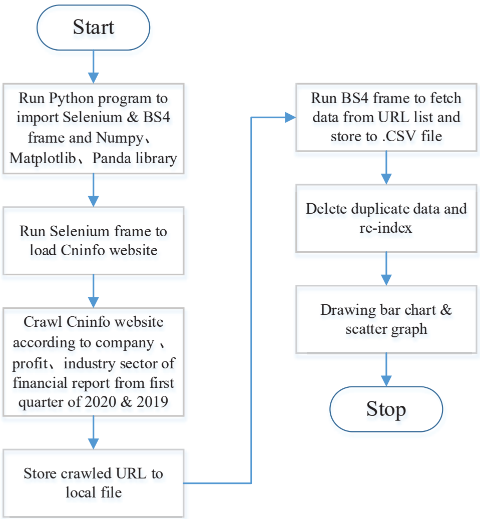
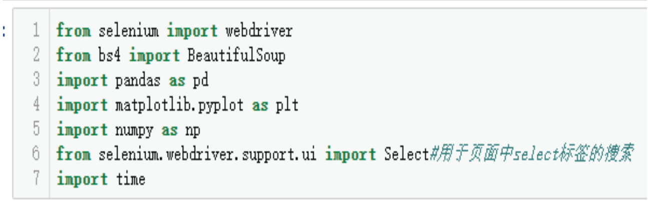

Start
Run Python program to import Selenium & BS4 frame and Numpy、Matplotlib、Panda library
Run Selenium frame to load Cninfo website
Crawl Cninfo website according to company、profit、industry sector of financial report from first quarter of 2020 & 2019
Store crawled URL to local file
Run BS4 frame to fetch data from URL list and store to .CSV file
Delete duplicate data and re-index
Drawing bar chart & scatter graph
Stop

Fig. 2. The process flow of website crawler & data visualization.

1) Get financial reportURL list

We will use Selenium to simulate the website operation, and Google webdriver to fetch the webpages source codes. Figure 3 is the program code of Python3.8 + Selenium + Chrome explorer.

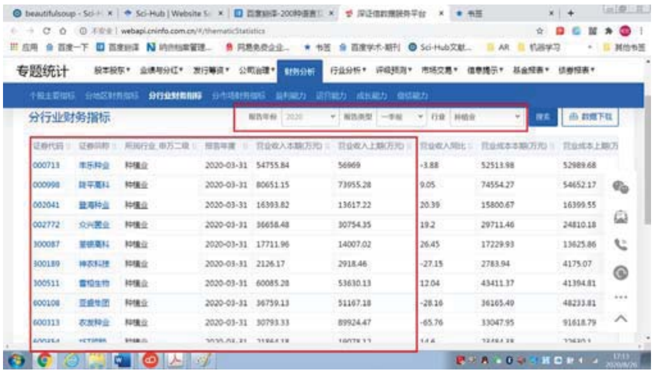


```
from selenium import webdriver
from bs4 import BeautifulSoup
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from selenium.webdriver.support.ui import Select#用于页面中select标签的搜索
import time
```

Fig. 3. Python Program code.

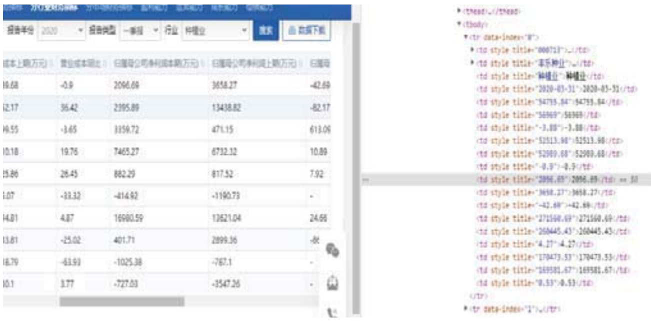
Figure 4 is the Cninfo website which we will crawl. In this website page, the net profit of each industry in the first quarter of 2019 and 2020 can be crawled, and then the stock code, company name and secondary industry category can be retrieved also (Figure 5).

Because the web page is a single page application, it uses selenium automatic positioning element. Therefore, it is necessary to simulate and click （Financial analysis)–> (Financial indicator by industry sector) by industry in the header of the page. In this page, companies in each industry need to be crawled. Therefore, we need selenium simulation button, click the next page, and then use selenium to simulate industry selection after crawling through an industry. Finally, simulate search by pressing button to continue the new climb. The taps which red underline marked in figure 6 is the indication of selenium click and select simulation.



证券代码	证券简称	所属行业_申万二级	报告年度	营业收入本期(万元)	营业收入上期(万元)	营业收入同比	营业成本本期(万元)	营业成本上期(万元)
000713	丰乐种业	种植业	2020-03-31	54755.84	56969	-3.88	52513.98	52989.68
000998	隆平高科	种植业	2020-03-31	80651.15	73955.28	9.05	74554.27	54652.17
002041	登海种业	种植业	2020-03-31	16393.82	13617.22	20.39	15800.67	16399.55
002772	众兴菌业	种植业	2020-03-31	36658.48	30754.35	19.2	29711.46	24810.18
300087	荃银高科	种植业	2020-03-31	17711.96	14007.02	26.45	17229.93	13625.86
300189	神农科技	种植业	2020-03-31	2126.17	2918.46	-27.15	2783.94	4175.07
300511	雪榕生物	种植业	2020-03-31	60085.28	53630.13	12.04	43411.37	41394.81
600108	亚盛集团	种植业	2020-03-31	36759.13	51167.18	-28.16	36165.49	48233.81
600313	农发种业	种植业	2020-03-31	30793.33	89924.47	-65.76	33047.95	91618.79

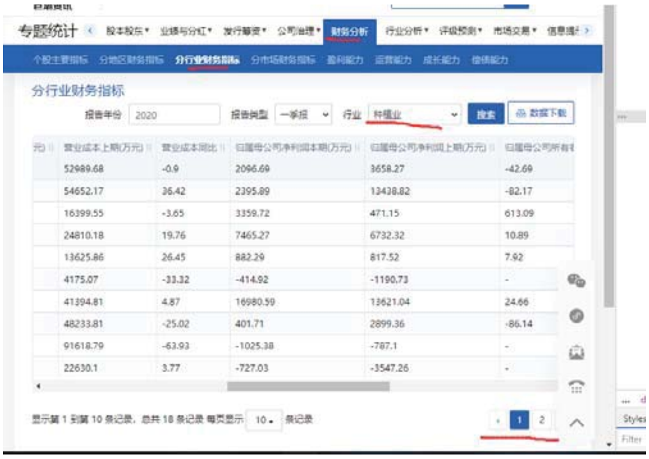
Fig. 4. Cninfo website with company’s income of current and last period.



本上期(万元)	营业成本同比	归属母公司净利润本期(万元)	归属母公司净利润上期(万元)	归属母
9.68	-0.9	2096.69	3658.27	-42.69
2.17	36.42	2395.89	13438.82	-82.17
9.55	-3.65	3359.72	471.15	613.09
0.18	19.76	7465.27	6732.32	10.89
5.86	26.45	882.29	817.52	7.92
.07	-33.32	-414.92	-1190.73	-
4.81	4.87	16980.59	13621.04	24.66
3.81	-25.02	401.71	2899.36	-8
8.79	-63.93	-1025.38	-787.1	-
0.1	3.77	-727.03	-3547.26	-

```
▸<thead>…</thead>
▾<tbody>
  ▾<tr data-index="0">
    ▸<td style title="000713">…</td>
    ▸<td style title="丰乐种业">…</td>
     <td style title="种植业">种植业</td>
     <td style title="2020-03-31">2020-03-31</td>
     <td style title="54755.84">54755.84</td>
     <td style title="56969">56969</td>
     <td style title="-3.88">-3.88</td>
     <td style title="52513.98">52513.98</td>
     <td style title="52989.68">52989.68</td>
     <td style title="-0.9">-0.9</td>
     <td style title="2096.69">2096.69</td> == $0
     <td style title="3658.27">3658.27</td>
     <td style title="-42.69">-42.69</td>
     <td style title="271560.69">271560.69</td>
     <td style title="260445.43">260445.43</td>
     <td style title="4.27">4.27</td>
     <td style title="170473.53">170473.53</td>
     <td style title="169581.67">169581.67</td>
     <td style title="0.53">0.53</td>
   </tr>
  ▸<tr data-index="1">…</tr>
```

Fig. 5. Company’S income of current and last period.

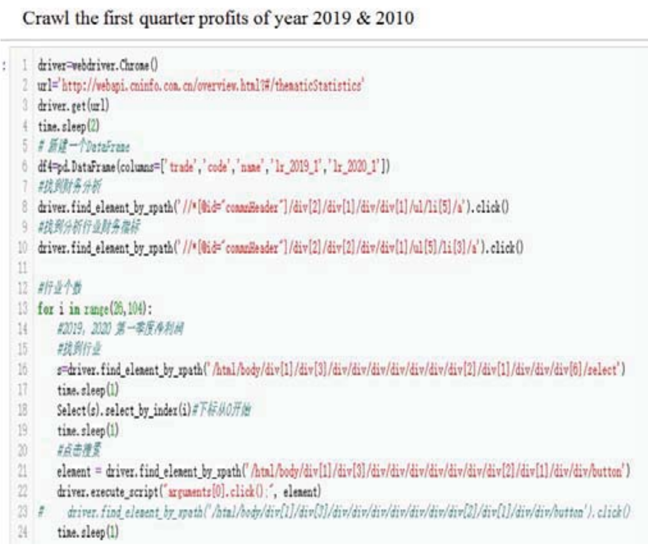


元)	营业成本上期(万元)	营业成本同比	归属母公司净利润本期(万元)	归属母公司净利润上期(万元)	归属母公司所有者
	52989.68	-0.9	2096.69	3658.27	-42.69
	54652.17	36.42	2395.89	13438.82	-82.17
	16399.55	-3.65	3359.72	471.15	613.09
	24810.18	19.76	7465.27	6732.32	10.89
	13625.86	26.45	882.29	817.52	7.92
	4175.07	-33.32	-414.92	-1190.73	-
	41394.81	4.87	16980.59	13621.04	24.66
	48233.81	-25.02	401.71	2899.36	-86.14
	91618.79	-63.93	-1025.38	-787.1	-
	22630.1	3.77	-727.03	-3547.26	-

显示第 1 到第 10 条记录，总共 18 条记录 每页显示 10 条记录

Fig. 6. The indication of selenium click and select simulation.

After crawling the URL lists for the first quarter data of 2019 & 2020, Python program will store those URL list in the Data Frame (df4) of Pandas. Figure 7 is part of source codes.



Crawl the first quarter profits of year 2019 & 2010

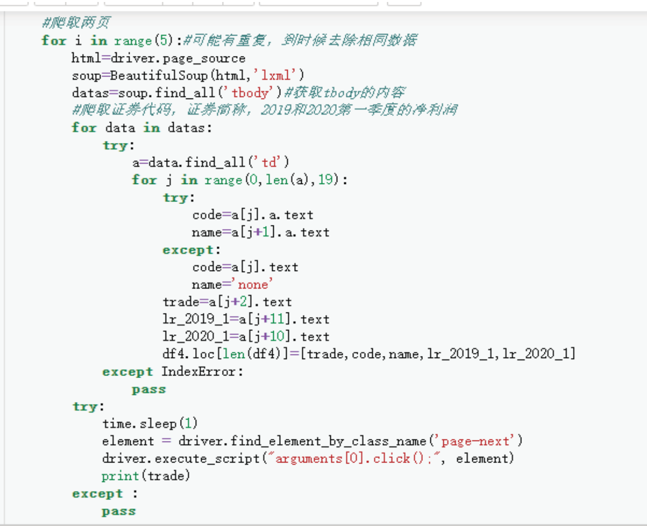
```
driver=webdriver.Chrome()
url='http://webapi.cninfo.com.cn/overview.html?#/thematicStatistics'
driver.get(url)
time.sleep(2)
# 新建一个DataFrame
df4=pd.DataFrame(columns=['trade','code','name','lr_2019_1','lr_2020_1'])
#找到财务分析
driver.find_element_by_xpath('//*[@id="commHeader"]/div[2]/div[1]/div/div[1]/ul/li[5]/a').click()
#找到分析行业财务指标
driver.find_element_by_xpath('//*[@id="commHeader"]/div[2]/div[2]/div/div[1]/ul[5]/li[3]/a').click()

#行业个数
for i in range(26,104):
    #2019, 2020 第一季度净利润
    #找到行业
    s=driver.find_element_by_xpath('/html/body/div[1]/div[3]/div/div/div/div/div/div/div[2]/div[1]/div/div/div[6]/select')
    time.sleep(1)
    Select(s).select_by_index(i)#下标从0开始
    time.sleep(1)
    #点击搜索
    element = driver.find_element_by_xpath('/html/body/div[1]/div[3]/div/div/div/div/div/div/div[2]/div[1]/div/div/button')
    driver.execute_script("arguments[0].click();", element)
#    driver.find_element_by_xpath('/html/body/div[1]/div[3]/div/div/div/div/div/div/div[2]/div[1]/div/div/button').click()
    time.sleep(1)
```

Fig. 7. Source code of crawling & store URL.

2) Get financil data from URL list

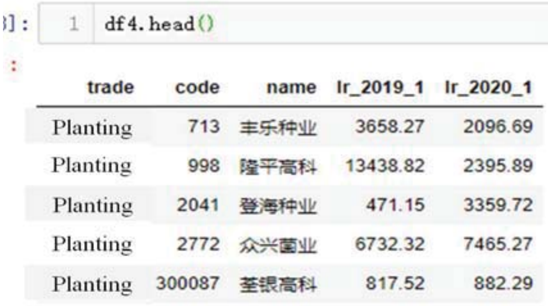
Next, we will use BeautifulSoup fetch and delete duplicated data embedded in the HTML or XML. Figure 8 is part of source codes.



```
#爬取两页
for i in range(5):#可能有重复，到时候去除相同数据
    html=driver.page_source
    soup=BeautifulSoup(html,'lxml')
    datas=soup.find_all('tbody')#获取tbody的内容
    #爬取证券代码，证券简称，2019和2020第一季度的净利润
    for data in datas:
        try:
            a=data.find_all('td')
            for j in range(0,len(a),19):
                try:
                    code=a[j].a.text
                    name=a[j+1].a.text
                except:
                    code=a[j].text
                    name='none'
                trade=a[j+2].text
                lr_2019_1=a[j+11].text
                lr_2020_1=a[j+10].text
                df4.loc[len(df4)]=[trade,code,name,lr_2019_1,lr_2020_1]
        except IndexError:
            pass
    try:
        time.sleep(1)
        element = driver.find_element_by_class_name('page-next')
        driver.execute_script("arguments[0].click();", element)
        print(trade)
    except :
        pass
```

Fig. 8. Source code of fetch and delete duplicate data.

Figure 9 is the example of crawled data which stored in df4.



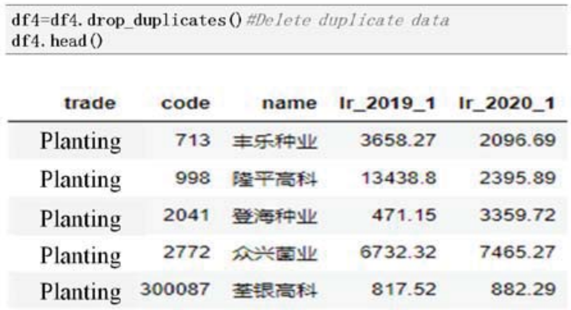
```
df4.head()
```

	trade	code	name	lr_2019_1	lr_2020_1
	Planting	713	丰乐种业	3658.27	2096.69
	Planting	998	隆平高科	13438.82	2395.89
	Planting	2041	登海种业	471.15	3359.72
	Planting	2772	众兴菌业	6732.32	7465.27
	Planting	300087	荃银高科	817.52	882.29

Fig. 9. Example of crawled data.

3) Pipes for storing crawling content

When crawling to the last web page, it will turn back to first web page again. Since the number of companies in each industry is different, so we should control the number of normal cycles. If climbing down df4, we can find out the re duplicate data. Therefore, we have to delete the duplicate data. Figure 10 is the Python program for processing duplicate data.

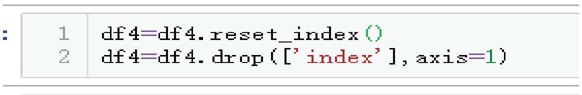


```
df4=df4.drop_duplicates() #Delete duplicate data
df4.head()
```

	trade	code	name	lr_2019_1	lr_2020_1
	Planting	713	丰乐种业	3658.27	2096.69
	Planting	998	隆平高科	13438.8	2395.89
	Planting	2041	登海种业	471.15	3359.72
	Planting	2772	众兴菌业	6732.32	7465.27
	Planting	300087	荃银高科	817.52	882.29

Fig. 10. Process duplicate data.

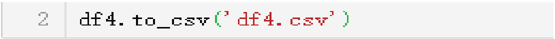
After deleting duplicate data, there will generate discontinuous indexes. Hence, we have to reset the indexes and drop original indexes. Figure 11 is the program of reset and drop indexes.



```
df4=df4.reset_index()
df4=df4.drop(['index'],axis=1)
```

Fig. 11. Program of reset and drop indexes.

Finally, the result will be converted to .csv file (df4.csv). See figure 12.

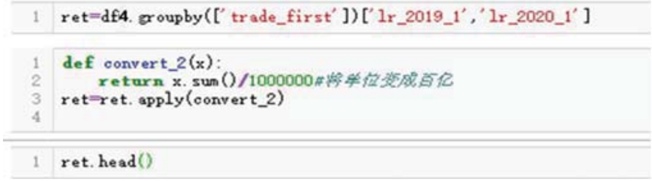


```
df4.to_csv('df4.csv')
```

Fig. 12. . Program of convert df4 to df4.csv file.

IV. Experiment Results

In this section, will use data visualization technology to analyze the data. The total net profit of each industry is obtained by grouping according to the first class industries, and then the original ten thousand dollars unit is changed into ten billion dollars unit. Figure 13 is the dollar unit convert program. Figure 14 is the result.



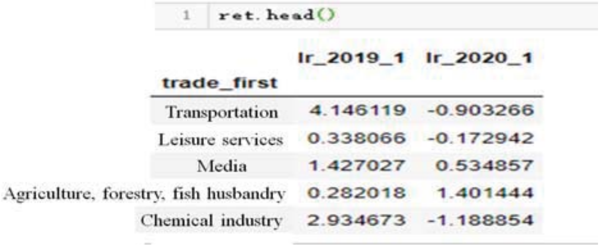
```
ret=df4.groupby(['trade_first'])['lr_2019_1','lr_2020_1']
```

```
def convert_2(x):
    return x.sum()/1000000#将单位变成百亿
ret=ret.apply(convert_2)

```

```
ret.head()
```

Fig. 13. Dollar unit converted program.



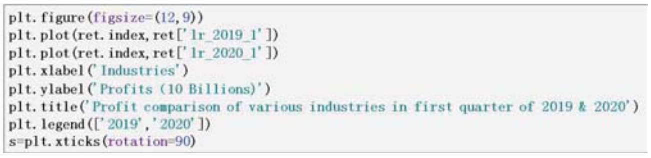
```
ret.head()
```

	lr_2019_1	lr_2020_1
trade_first		
Transportation	4.146119	-0.903266
Leisure services	0.338066	-0.172942
Media	1.427027	0.534857
Agriculture, forestry, fish husbandry	0.282018	1.401444
Chemical industry	2.934673	-1.188854

Fig. 14. Dollar unit converted result.

A. Bar Chart of Industries' Profit

Figure 15 is the bar chart drawing program. Figure 16 is the bar chart.



```
plt.figure(figsize=(12,9))
plt.plot(ret.index,ret['lr_2019_1'])
plt.plot(ret.index,ret['lr_2020_1'])
plt.xlabel('Industries')
plt.ylabel('Profits (10 Billions)')
plt.title('Profit comparison of various industries in first quarter of 2019 & 2020')
plt.legend(['2019','2020'])
s=plt.xticks(rotation=90)
```

Fig. 15. Bar chart drawing program.

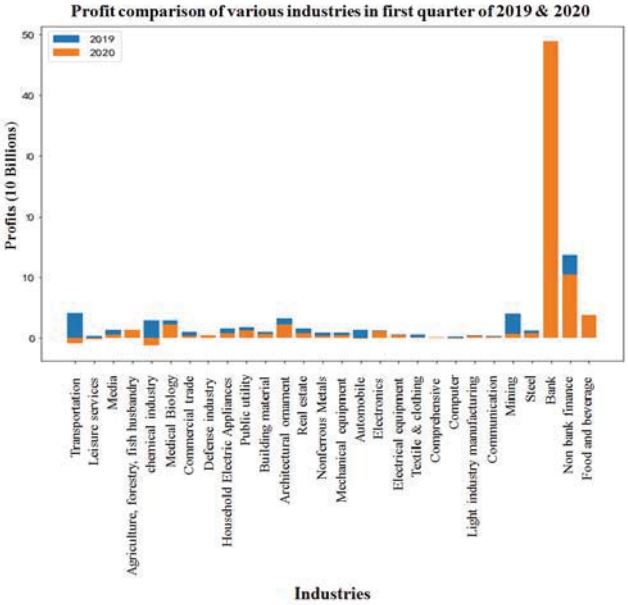
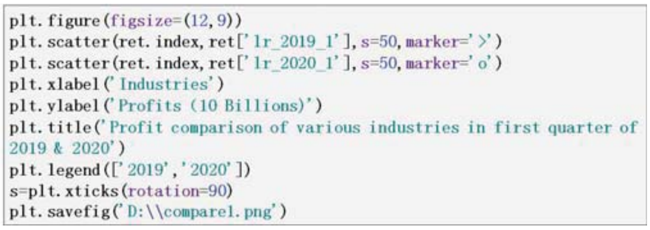


Fig. 16. Bar chart of first quarter profit of 2019 & 2020.

It can be seen that in 2020, except for agriculture, forestry, animal husbandry and fish industry, national defense army industry, banking and food and beverage industry, almost all industries have depressed due to the epidemic situation. Because in this histogram, the first quarter of 2020 presents at the upper level, and the first quarter of 2019 presents at the lower level. When the data of the first quarter of 2020 is greater than that of the first quarter of 2019 for the same industry, the data histogram of the first quarter of 2019 cannot be seen. Therefore, we use the scatter plot to present.

B. Scatter Plot of Industries' Profit

Figure 17 is the scatter plot drawing program. Figure 18 is the scatter plot.



```
plt.figure(figsize=(12,9))
plt.scatter(ret.index,ret['lr_2019_1'],s=50,marker='>')
plt.scatter(ret.index,ret['lr_2020_1'],s=50,marker='o')
plt.xlabel('Industries')
plt.ylabel('Profits (10 Billions)')
plt.title('Profit comparison of various industries in first quarter of
2019 & 2020')
plt.legend(['2019','2020'])
s=plt.xticks(rotation=90)
plt.savefig('D:\\compare1.png')
```

Fig. 17. Scatter plot drawing program.

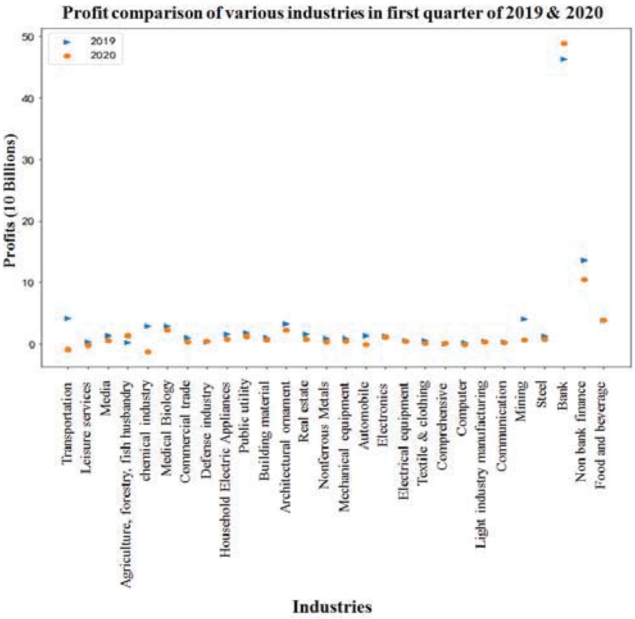
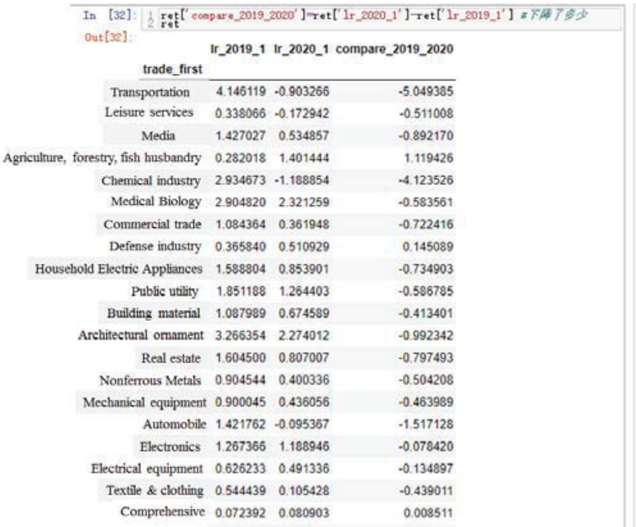


Fig. 18. Scatter plot of first quarter profit of 2019 & 2020.

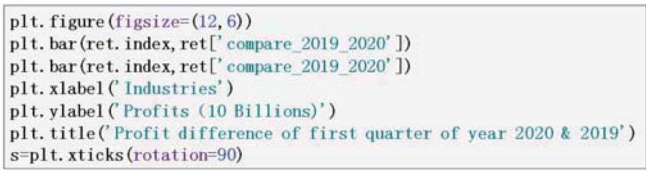
C. Bar Chart of Profit Difference

In order to compare the decline of each industry in the first quarter, we use the net profit of the first quarter of 2020 to minus the net profit of the first quarter of 2019 and get the following result. Figure 19 is the program and result. Figure 20 is the bar chart drawing program and figure 21 is the result.



```
In  [32]:  ret['compare_2019_2020']=ret['lr_2020_1']-ret['lr_2019_1'] #下降了多少
           ret
Out[32]:
                                      lr_2019_1  lr_2020_1  compare_2019_2020
                         trade_first
                      Transportation   4.146119  -0.903266          -5.049385
                    Leisure services   0.338066  -0.172942          -0.511008
                               Media   1.427027   0.534857          -0.892170
Agriculture, forestry, fish husbandry   0.282018   1.401444           1.119426
                   Chemical industry   2.934673  -1.188854          -4.123526
                     Medical Biology   2.904820   2.321259          -0.583561
                    Commercial trade   1.084364   0.361948          -0.722416
                    Defense industry   0.365840   0.510929           0.145089
       Household Electric Appliances   1.588804   0.853901          -0.734903
                      Public utility   1.851188   1.264403          -0.586785
                   Building material   1.087989   0.674589          -0.413401
              Architectural ornament   3.266354   2.274012          -0.992342
                         Real estate   1.604500   0.807007          -0.797493
                   Nonferrous Metals   0.904544   0.400336          -0.504208
                Mechanical equipment   0.900045   0.436056          -0.463989
                          Automobile   1.421762  -0.095367          -1.517128
                         Electronics   1.267366   1.188946          -0.078420
                Electrical equipment   0.626233   0.491336          -0.134897
                   Textile & clothing   0.544439   0.105428          -0.439011
                       Comprehensive   0.072392   0.080903           0.008511
```

Fig. 19. Profit difference of program and result.



```
plt.figure(figsize=(12,6))
plt.bar(ret.index,ret['compare_2019_2020'])
plt.bar(ret.index,ret['compare_2019_2020'])
plt.xlabel('Industries')
plt.ylabel('Profits (10 Billions)')
plt.title('Profit difference of first quarter of year 2020 & 2019')
s=plt.xticks(rotation=90)
```

Fig. 20. Bar chart drawing program for profit difference.

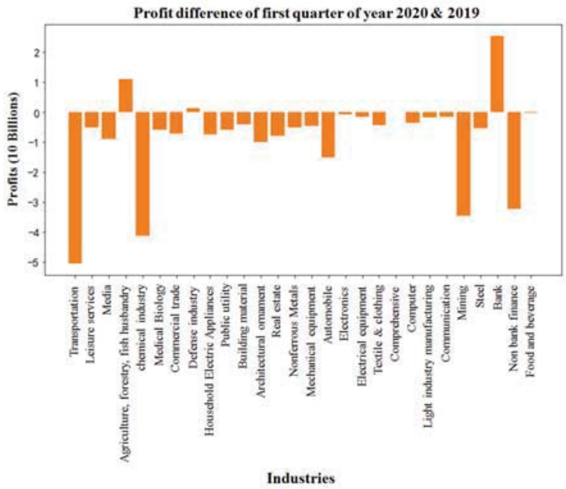


Fig. 21. Profit difference of first quarter of 2020 & 2019.

V. Conclusion

In this paper, we use web crawl technology and data visualization technology to crawl data from Cninfo website (http://www.cninfo.com.cn/new/index). From figure 21, we can see that most of industries' profits have declined by comparing the first quarter of 2020 and 2019 except for agriculture, forestry, animal husbandry and fish industry, national defense army industry, banking, and food and beverage industry. Among them, transportation, chemical industry, mining, non-banking and financial sector decreased the most. Because governments lockdown of the city from all outside contact and people have to stay at home. However, people still need to eat and purchase stuff from the internet. Government also invests money to rescue crisis enterprises. Therefore, banking, food and beverage industry still grows. On the whole, various industries have been hit by the COVID-19 epidemic.

References

[1] C. Huang, Y. Wang, and X. Li, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Jan 24. pii: S0140-6736(20)30183-5. doi: 10.1016/S0140-6736(20)30183-5. [Epub ahead of print]

[2] N. Zhu, D. Zhang, and W. Wang, A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Jan 24. doi: 10.1056/NEJMoa2001017. [Epub ahead of print]

[3] Q. Li, X. Guan, and P. Wu, Early transmission dynamics in Wuhan, China, of novel coronavirus-Infected pneumonia. N Engl J Med 2020 Jan 29.doi:10.1056/NEJMoa2001316 [Epub ahead of print]

[4] World Health Organization. WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020 at https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020. Published February 11, 2020.

[5] Worldometer: COVID-19 CORONAVIRUS PANDEMIC. https://www.worldometers.info/coronavirus/.

[6] https://www.brecorder.com/2020/03/20/581747/what-economic-impact-will-covid-19-bring-to-china/.

[7] D. Feisong and L. Quanjin, "Network data acquisition and classification in big data environment," Light Ind Technol. 2019; vol. 35, no. 7, pp. 68-71.

[8] Z. Jian and L. Qin, "The application of big data network crawler technology for architectural culture and environment protection," Concurrency Computat Pract Exper. 2020;e5769. https://doi.org/10.1002/cpe.5769

[9] L. Yang, "Design and Implementation of Distributed Web Crawler System Based on Marker Template," Wuhan, China: Huazhong University of Science and Technology; 2019.

[10] M. Russo, L. Carnevali, and V. Russo, "Modeling and deterioration mapping of facades in historical urban context by close-range ultra-lightweight UAVs photogrammetry," Int J Archit Herit. vol. 8, pp. 1-20, 2018.

[11] F. Ying and Z. Zhang, "Data Visualization Analysis of Big Data Recruitment Positions in Hangzhou Based on Python", Review of Computer Engineering Studies, Vol. 6, No. 4, December, 2019, pp. 81-86.

[12] M. Du and X. Yuan, "A survey of competitive sports data visualization and visual analysis," The Visualization Society of Japan 2020, Accepted: 6 May 2020.

[13] B. García, M. Gallego, F. Gortázar, and M. Munoz-Organero, "A Survey of the Selenium Ecosystem," Electronics 2020, vol.9, pp.1-29, 1067; doi:10.3390/electronics9071067 www.mdpi.com.

[14] SeleniumHQ Browser Automation "http://www.seleniumhq.org".

[15] Gábor László Hajba, Website Scraping with Python: Using BeautifulSoup and Scrapy, Aprress, Sopron, Hungary, ISBN-13 (pbk): 978-1-4842-3924-7, ISBN-13 (electronic): 978-1-4842-3925-4, https://doi.org/10.1007/978-1-4842-3925-4

[16] Chengfang Shenyang, Dalong Mo。Application and Using Tips of Beautifulsoup Database in Web Crawlers. Computer Knowledge and Technology,Vol.15, No.28, October. 2019. ISSN 1009-3044

[17] J. Hunt, Advanced Guide to Python3 Programming. Undergraduate Topics in Computer Science, pp. 35-42. © Springer Nature Switzerland AG 2019. ISBN 978-3-030-25942-6. ISBN 978-3-030-25942-6 (ebook).

[18] M. Bauer and M. Garland, Legate NumPy: accelerated and distributed array computing, Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, November 2019, Article No.: 23 pp. 1–23. https://doi.org/10.1145/3295500.3356175

[19] P. Lemenkova, Processing oceanographic data by Python libraries NumPy, SciPy and Pandas. Aquatic Research, ScientificWebJournals, 2019, Vol. 2, No. 2, pp.73-91.