TASK 4 - FACTIVA GUIDE TO OBTAIN COMPANY CODE & FREQUENCY

Aug 2021

OUR PURPOSE:

- Search for the frequency that a company is mentioned in a newspaper in each quarter in the period January 1995 – June 2020
- We follow these steps:
  1. Find the associated Factiva company code, industry information…
  2. Use the newspaper/source code and Factiva company code to search for the frequency that the company is mentioned during 01/01/1995-06/30/2020 in a newspaper.
  3. Repeat step 2 for other top newspapers (we have 11 newspapers).
  4. Repeat step 2 and 3 for other Factiva company codes.

Contents

I. RETRIEVE THE PART OF COMPANIES YOU WILL WORK ON 1

II. OBTAIN FACTIVA COMPANY CODE & SECTOR CODE 1

III. OBTAIN FREQUENCY 6

IV. UPLOAD RESULTS AND WEEKLY REPORT 11

V. DUPLICATES FACTIVA_COMPANY_NAME1 12

VI. SPECIAL CASE OF COMPANY NAME 13

VII. MISSING FCODEINCODE AND COMPANY INFORMATION 14

I. RETRIEVE THE PART OF COMPANIES YOU WILL WORK ON

- We have the list of company names obtained from STEP 1 in file "Factiva project\Factiva guide\STEP4_Factiva_company_list.xlsx" – column "Factiva_company_name" or "Factiva_company_name1".
- use R code or Python code (STEP4_Company_frequency.R or STEP4_Company_frequency.py) to retrieve the part of companies from file "STEP4_Factiva_company_list.xlsx", the code will export that part of data as .xlsx file
- Note: Your file should be named in the format: **"STEP4_Factiva_company_list_No<NUMBER 1>_<NUMBER 2>_YOURNAME.xlsx"**

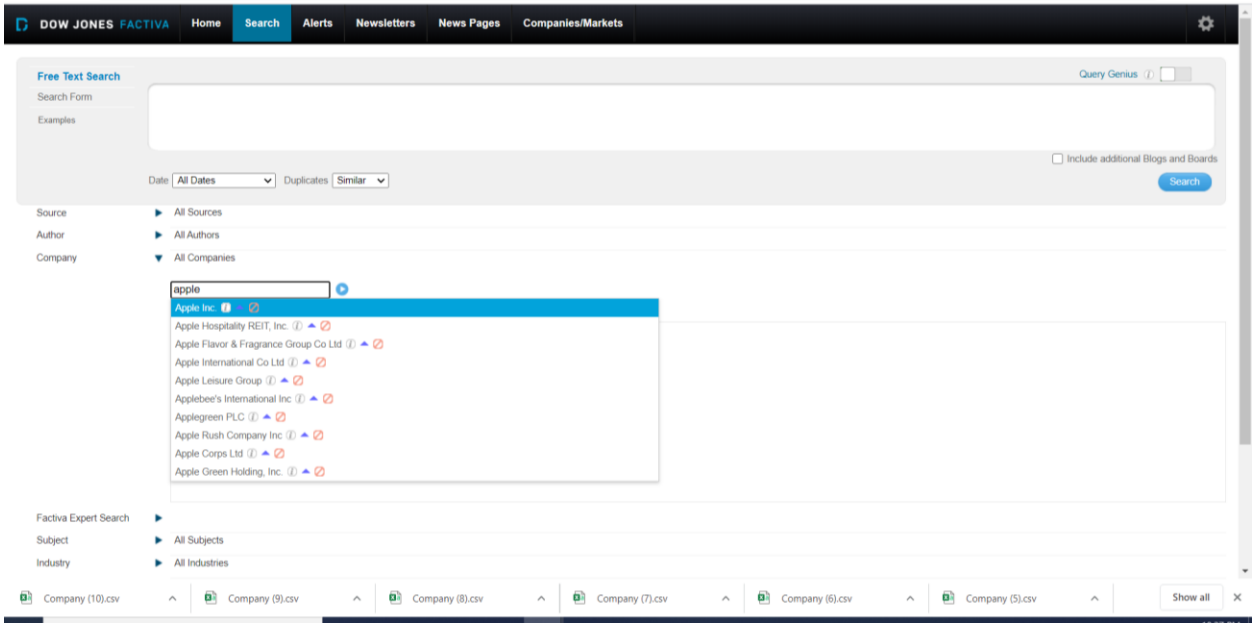
where <NUMBER 1> is the first company ID you will do and <NUMBER 2> is the last company ID you will do.

For example: "**STEP4_Factiva_company_list_No1_20_Ha.xlsx**": means Ha is doing companies from ID 1 to 20.

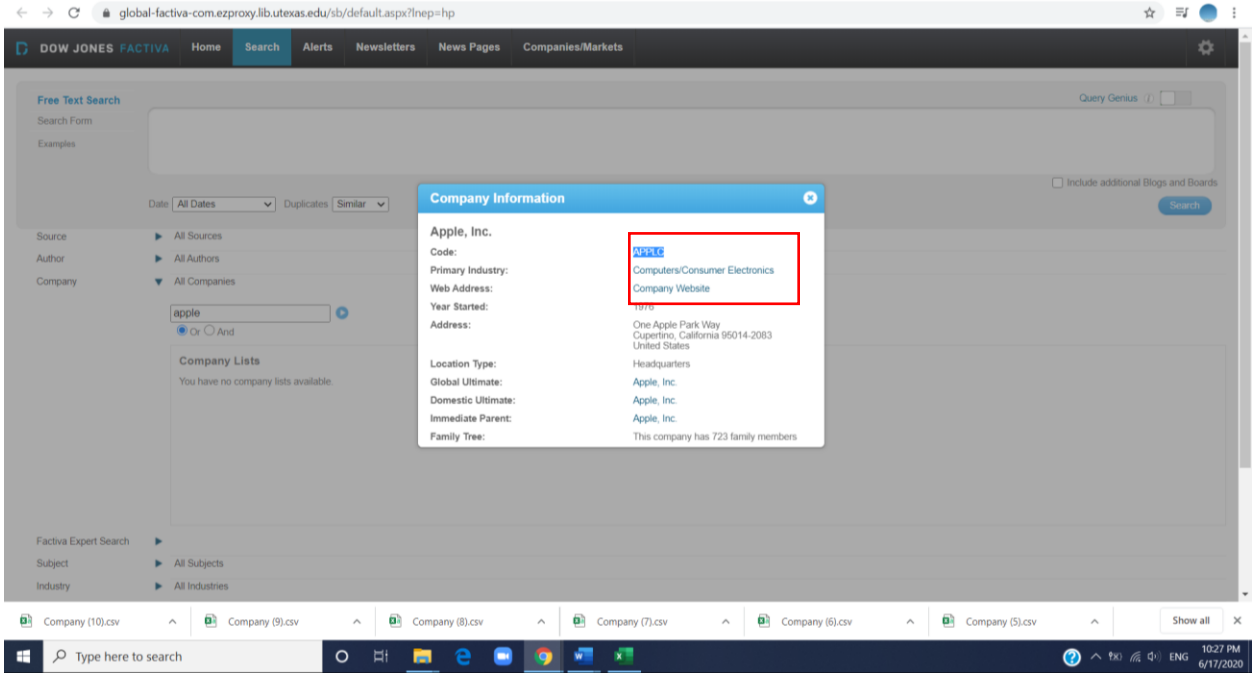
II. OBTAIN FACTIVA COMPANY CODE & SECTOR CODE

The steps below are to obtain the associated Factiva company code, the industry/sector and other information of the company. You will use the file you just extracted:

Step 1: From Factiva, put the name of the company into the search box, click the icon ⓘ next to the name of the company



Step 2: The information window will be opened, and we can see the Factiva company code associated with Apple Inc is APPLC. Then copy-paste this code into your file (For example "**STEP4_Factiva_company_list_No1_20_Ha.xlsx"** – column "factiva_company_code")



Step 3: If it is an organization instead of a company, then type 0 into the column "company_flag" (type 1 if it is a company). For example, European Union is an organization as we do not see information for "primary industry".

Step 4: Copy-paste the following information into the columns:

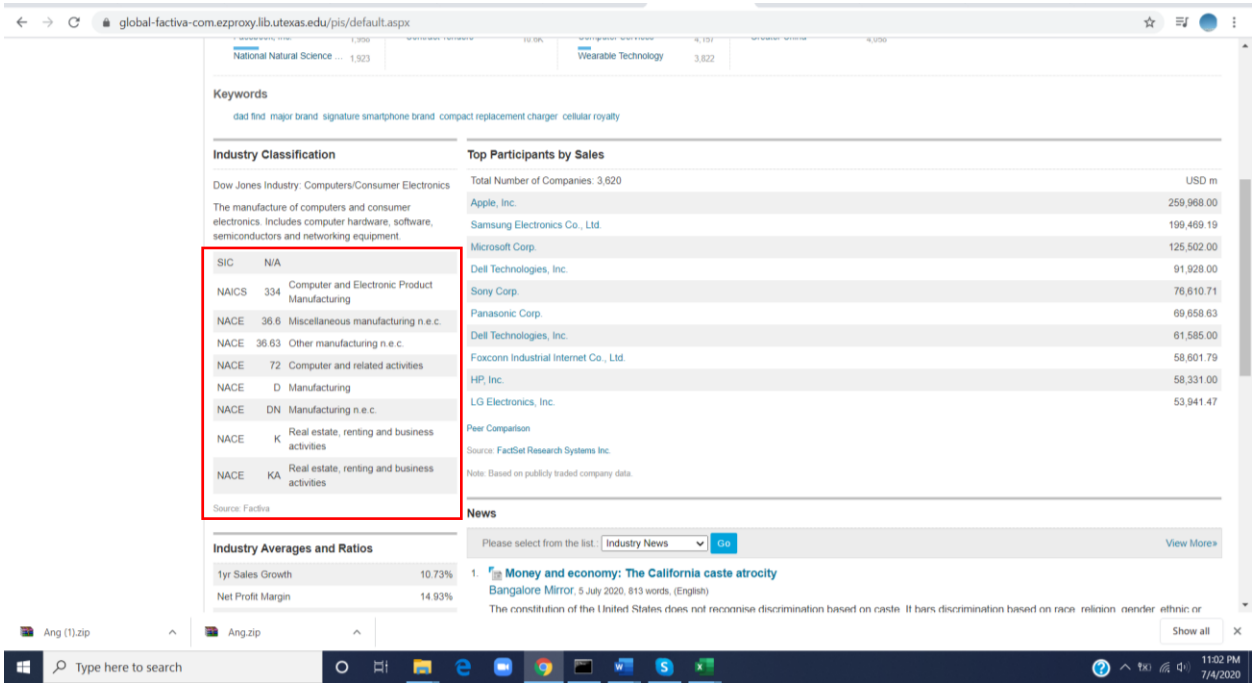
- "Primary industry" into the column "factiva industry".
  For example: copy-paste "Computers/Consumer Electronics" into the column "factiva industry"
- "address" into the column "address"
- "Location type" into the column "location_type"
- "Year started" into the column "year_started"
- "Global Ultimate" into the column 'global_ultimate"

- "Domestic Ultimate" into the column "domestic_ultimate"
- "Immediate Parent" into the column "immediate_parent"

For example:

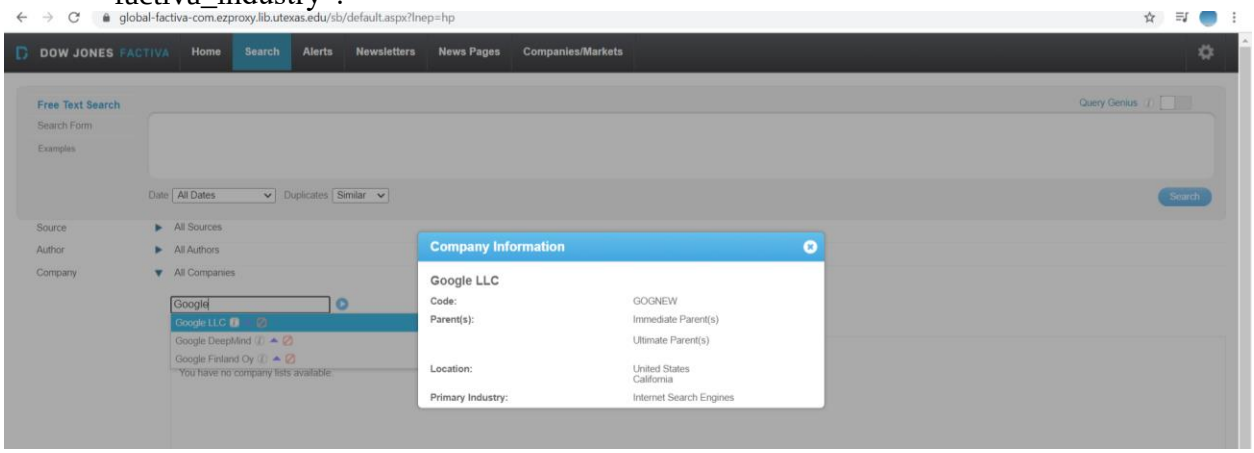
Step 5: Write the country of origin of the company into column "company_country", for example, here it is "US". You can find 2-letter country code list in File "STEP4_Company_frequency.xlsx"-sheet "country_code" (it is ISO ALPHA-2)

Step 6: Click the primary industry, a new page will be opened. Copy-paste the NAICS code into the column "NAICS Subsector", and the description of that NAICS Subsector into the column "NAICS Subsector Description". For example: copy-paste "334" into the column "NAICS Subsector", and "Computer and Electronic Product Manufacturing" into the column "NAICS Subsector Description". Similar work for NACE/SIC code.

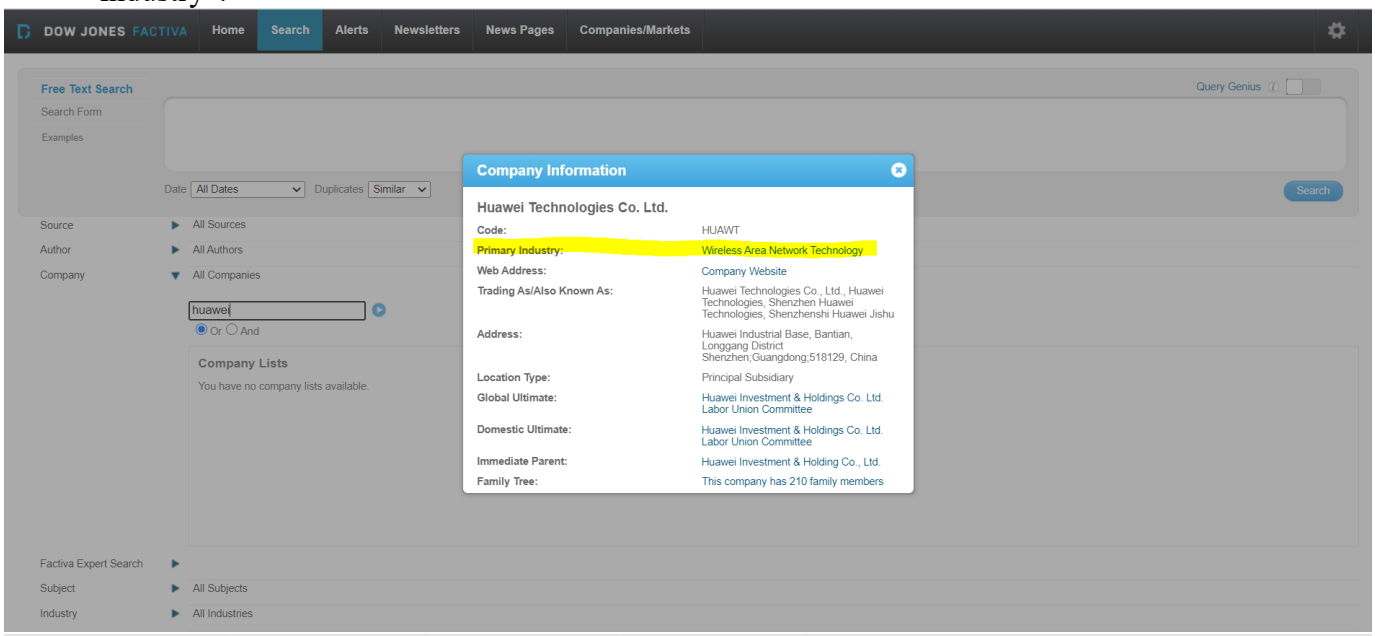


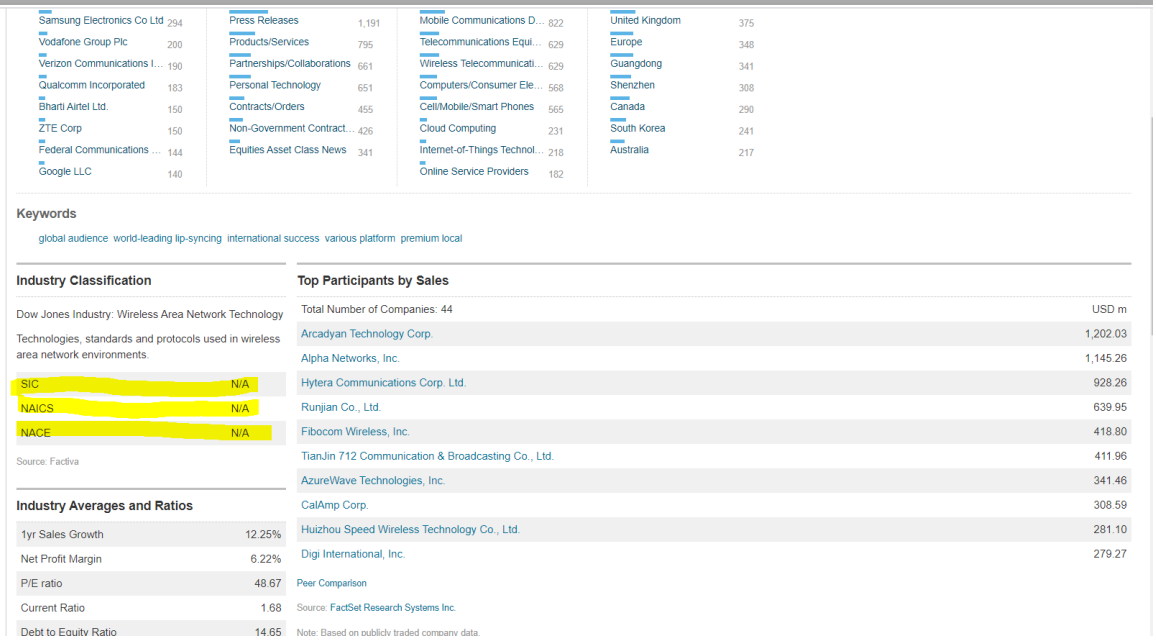
NOTE:

- It is possible that we could not be able to click the "primary industry" to access NAICS/SIC/NACE code like the example below. In this case, just copy-paste "Internet Search Engines" into the column "factiva_industry".

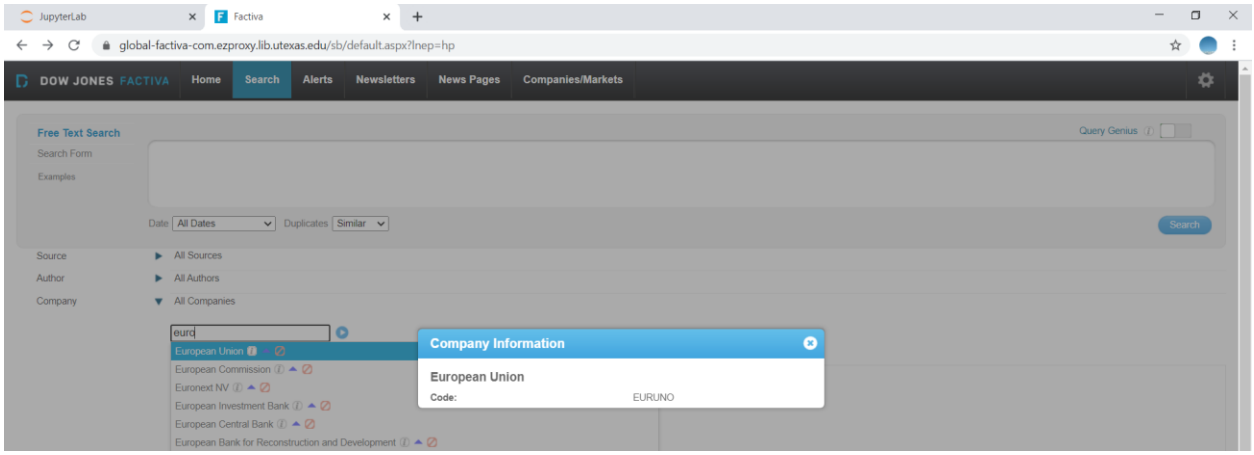


- It is possible that Factiva does not have NAICS/SIC/NACE code of the industry like the example below. In this case just copy-paste “Wireless Area Network Technology” into the column “factiva industry”.

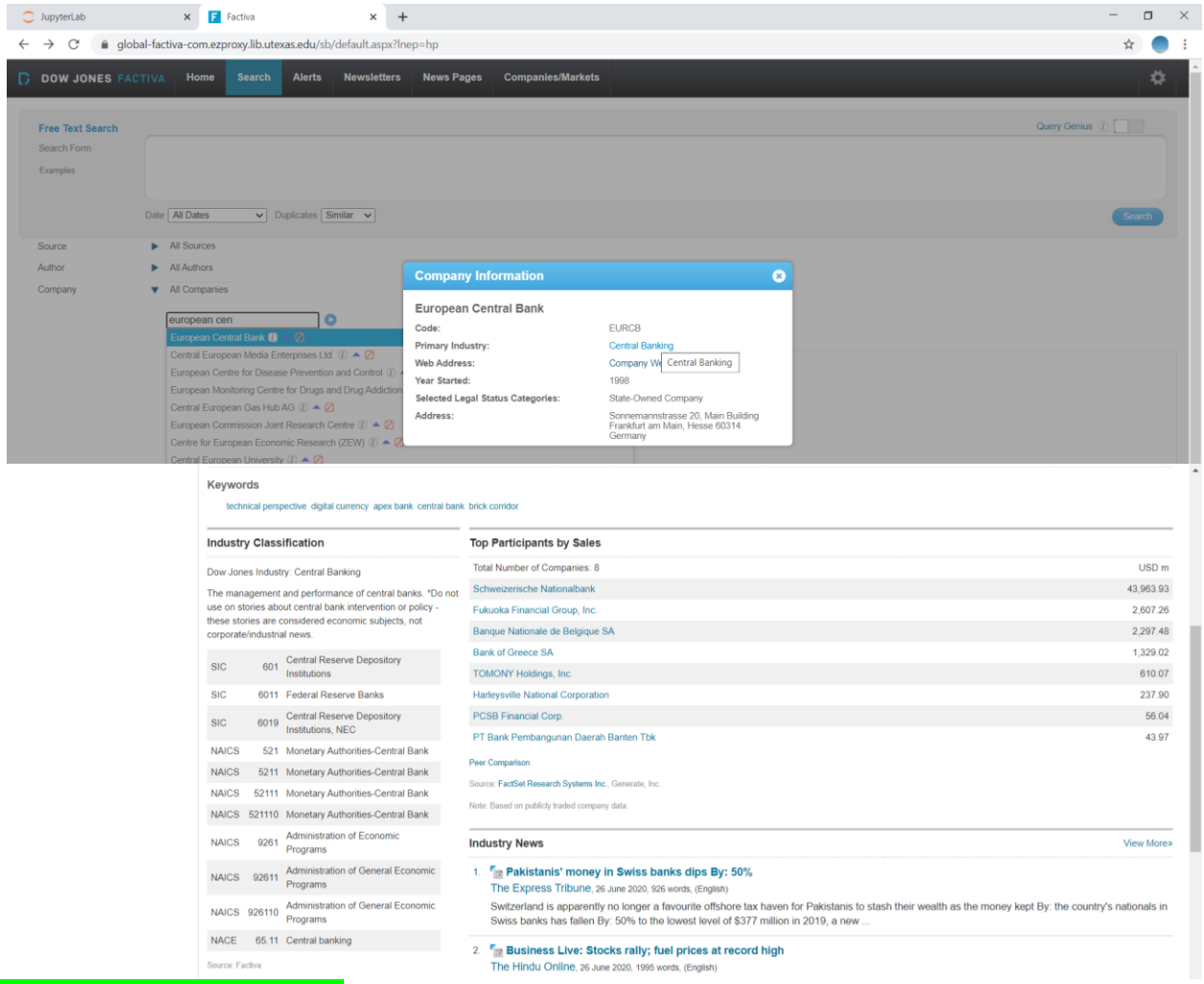




- It is possible that Factiva gives us both NAICS/SIC, NACE code. Then look into the address of the company, if it is in the US then NAICS/SIC is the most important information, if it is in other countries then NACE/SIC is the most important information (the example of Apple Inc. – code APPLC above). If you are not sure whether you should get NAICS/SIC/NACE code, then get all possible information available.
- If it is an organization instead of a company, then type 0 into the column “Company” (type 1 if it is a company). For example, European Union is an organization as we do not see information for “primary industry”



- It is possible that Factiva gives us multiple codes for NAICS/SIC/NACE, then take the most general code of each of them. For example, European Central Bank
  Then copy paste:
  - "601" and "Central Reserve Depository Institutions" into column "SIC subsector" and "SIC subsector description", then
  - "521" and "Monetary Authorities-Central Bank" into column "NAICS subsector" and "NAICS subsector description", then
  - "65.11" and "Central banking" into column "NACE subsector" and "NACE subsector description".



Step 7: column "fcodeincode".

- use R/Python code to merge (file STEP4_Company_frequency.R or STEP4_Company_frequency.py)

⇨ Final product of this work is a file named "STEP4_Factiva_company_list_No<NUMBER 1>_<NUMBER 2>_sector_YOURNAME.xlsx"
For example: "**STEP4_Factiva_company_list_No1_20_sector_Ha.xlsx**":

III. OBTAIN FREQUENCY

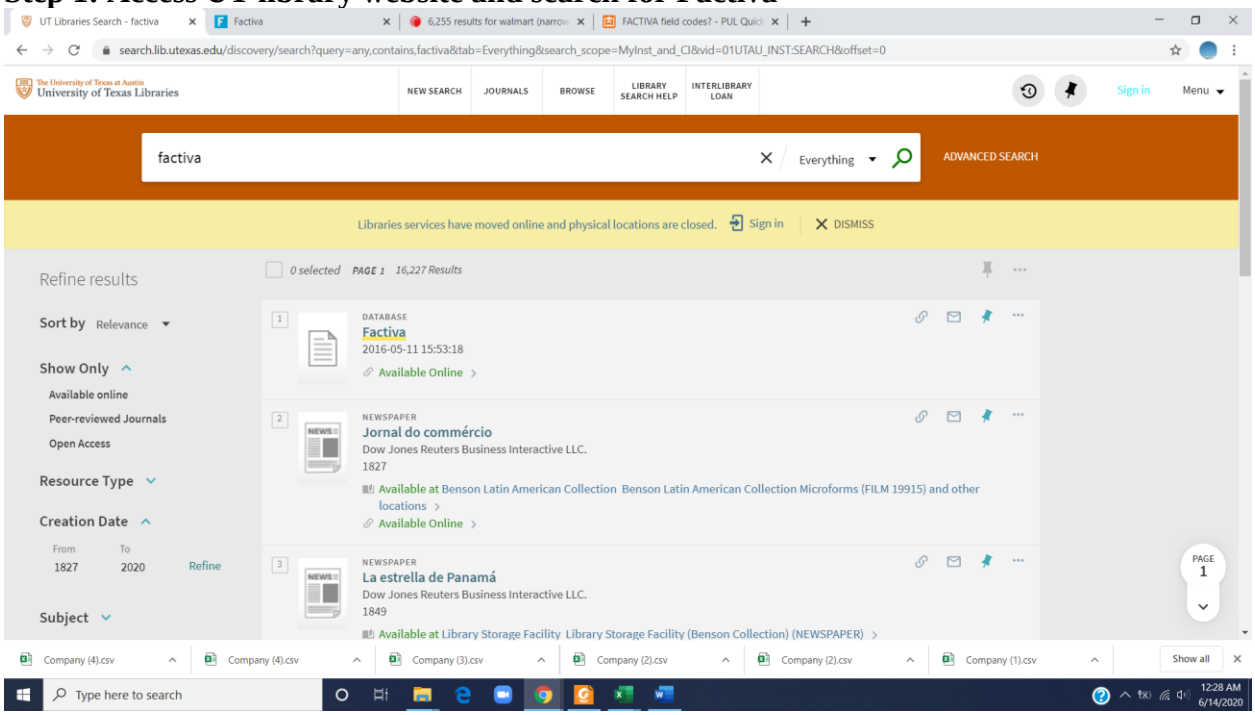
The steps below are to obtain the frequency that a company name is mentioned in a newspaper in a quarter in the period 01/01/1995-06/30/2020 in Factiva. Below is an example of how to get the frequency.

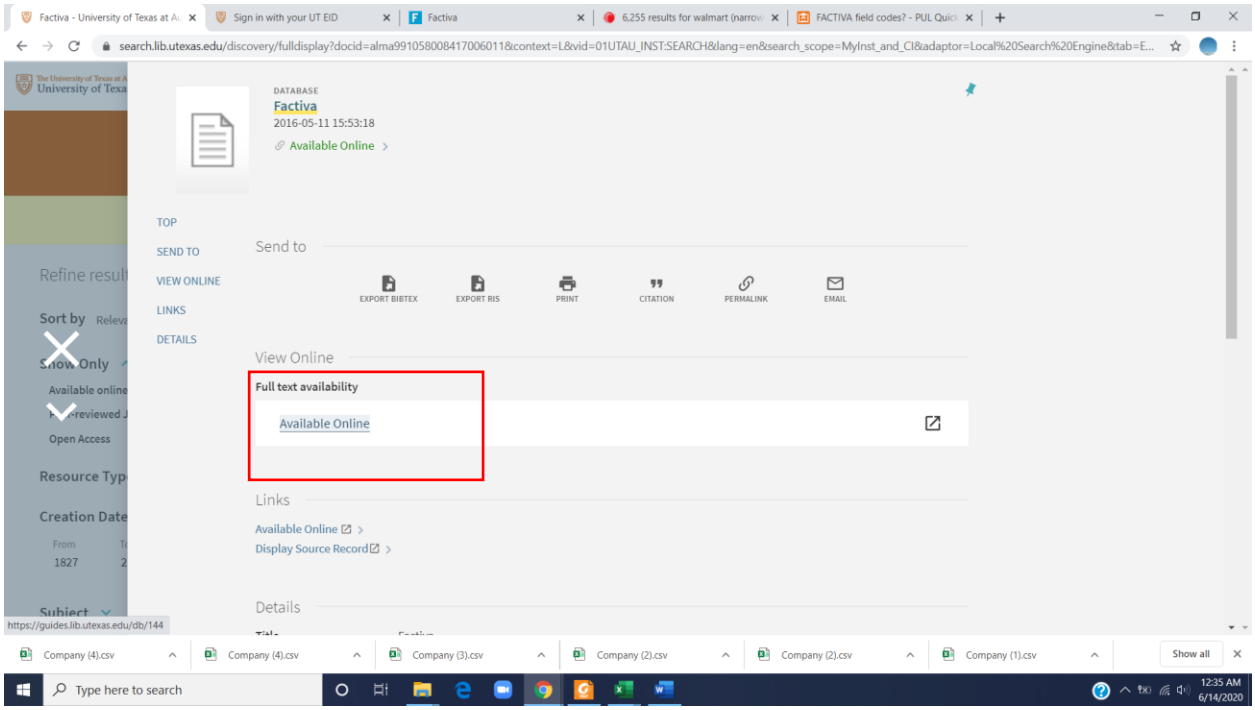
- Your file's name format is: STEP4_Company_frequency_No<NUMBER 1>_<NUMBER 2>_YOURNAME.xlsx
  For example: STEP4_Company_frequency_No1_20_Ha.xlsx
- You can use file "STEP4_Company_frequency_template.xlsx" as a template.
- Please put all companies in 1 sheet. Please don't create each sheet for each company.

NOTE:

- We only need to obtain quarterly frequency for "company", no need to do for "organization" (organization examples: European Union...).
- In file "STEP4_Factiva_company_list.xlsx": all observations with _merge1 ==3 are companies, so we will focus on doing them first.

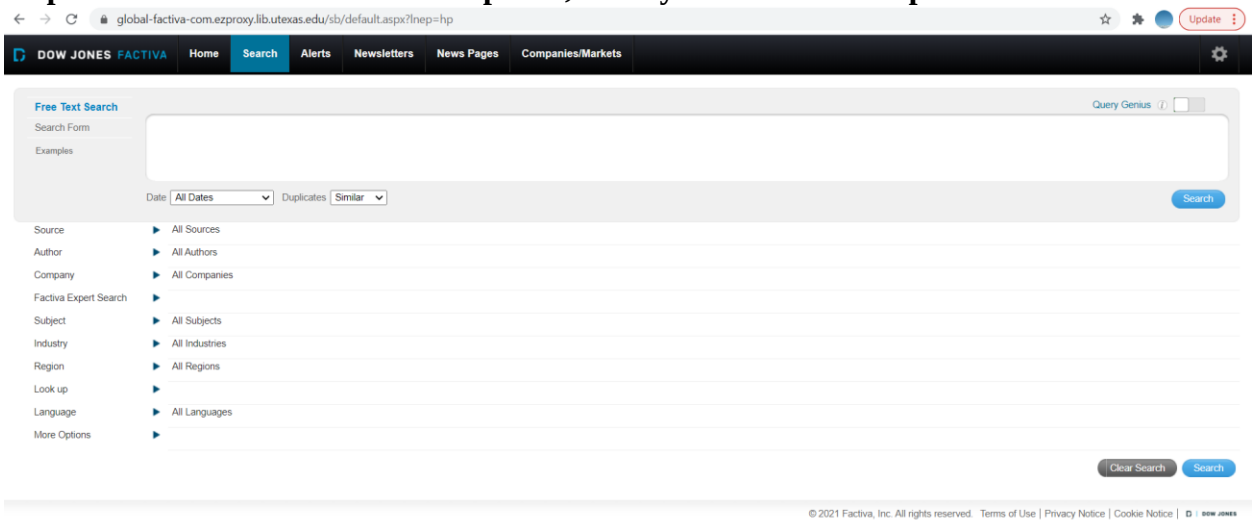
Step 1: Access UT library website and search for Factiva





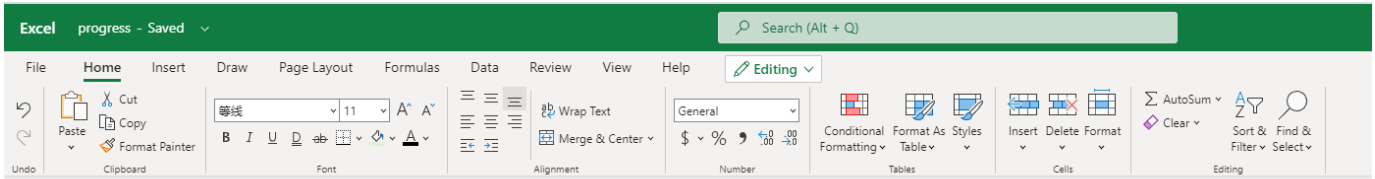
Click on "Available Online" to access Factiva.

Step 2: A new window/tab will be opened, enter your UT EID and password to access Factiva



Step 3: Perform the search with following steps:

1. Set Date: 01/01/1995 to 06/30/2020
2. Search syntax: (rst=…OR rst=… OR rst=…) AND fds=…
   where "rst" means "source code" of the newspaper, "fds" is Factiva company code.
   You can find all search syntaxes in file "progress.xlsx" or file "STEP4_Company_frequency.xlsx"



	A	B	C	D	E
1	Newspaper_cour	Newspaper_code	Newspaper_name	search command	
2	US	sfusat	USA Today (All sources)	rst=sfusat and fds=….	
3	US	sfwsj	The Wall Street Journal (All Sources)	rst=sfwsj and fds=….	
4	US	sfnyt	The New York Times (All Sources)	rst=sfnyt and fds=….	
5	CA	sfglob	The Globe and Mail (All sources)	rst=sfglob and fds=….	
6	GB	sfftftp	Financial Times (All sources)	(rst=ftft or rst=ftcom) and fds=….	
7	DE	sfsddzp	Suddeutsche Zeitung (All sources)	(rst=sddz or rst=sudzeit or rst=sz) and fds=….	
8	IT	sfcordesp	Corriere della Sera (All sources)	(rst=cordes or rst=coronl or rst=corsup) and fds=….	
9	ES	sfelpaisp	El Pais (All sources)	(rst=paisba or rst=paisbi or rst=paise or rst=paisma or rst=paisn or rst=psalud or rst=paisse or rst=paisva or rst=ep or rst=elpais) and fds=…	
10	FR	sffigarop	Le Figaro (All sources)	(rst=lf or rst=figaro or rst=figbour or rst=figetu or rst=figmag or rst=figpre) and fds=…	
11	JP	sfmaip	Mainichi Shimbun (All sources)	(rst=mainhq or rst=mainlo or rst=maimep or rst=maicnr or rst=maioti) and fds=...	
12	JP	sfsankeip	Sankei Shimbun (All sources)	(rst=snkosa or rst=snkeis or rst=snkspo) and fds=...	
13					

3. Duplicates: Identical
4. Language: All languages

➔ Get the frequency and enter that number into 'frequency' column in excel file.

Note:

- The bracket (…..) is extremely important when working on a newspaper which consists of multiple source codes. If we forget it, we will get wrong number for frequency.
- Setting Language to be "All languages" are extremely important for newspapers written in languages other than English.
- Setting Duplicates to be 'Identical' is extremely important to eliminate duplicates in frequency.
- Frequency = Number of headlines – total duplicates (until the last page). When will we see frequency = 0?
- If we see the result of the search are multiple pages, we have to go to the last page to get the correct number for frequency.

Step 4: column "small_company": it is a flag

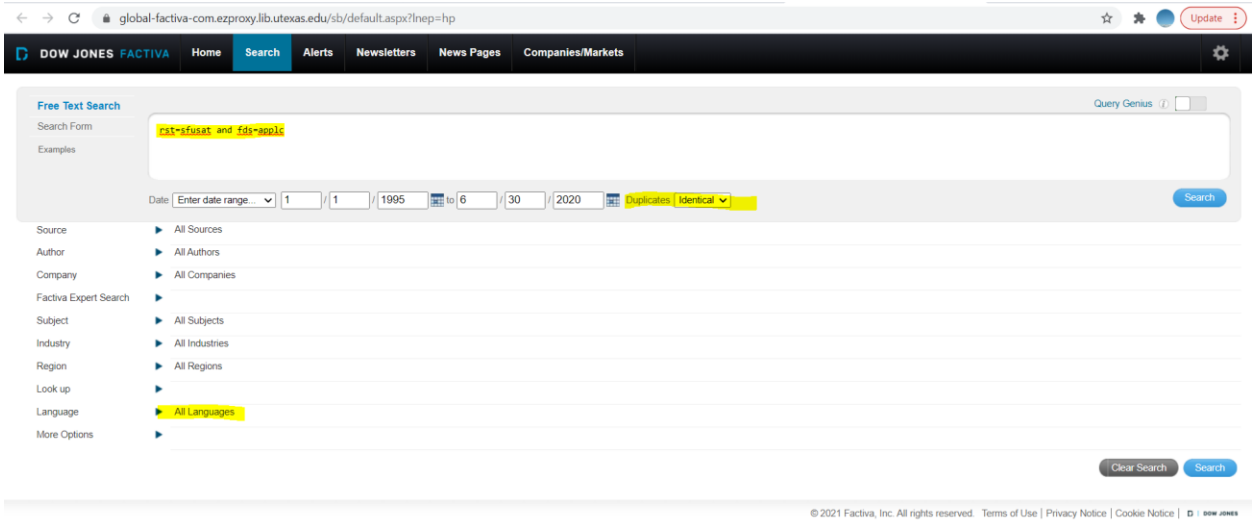
- If the name is a COMPANY and it is NOT SMALL: put 0 into this column.
- If the name is a COMPANY and it is SMALL: put 1 into this column.
- If the name is NOT a COMPANY (but an organization) and it is NOT SMALL: put 2 into this column.
- If the name is NOT a COMPANY (but an organization) and it is SMALL: put 3 into this column.

Small company definition: have frequency = 0 for the whole period of 01/01/1995-06/30/2020.
This flag can be different for same company in different newspapers. For example, a company could have 0 frequency in sfmaip (small_company is 1), but has positive frequency in sfwsj (small_company is 0)

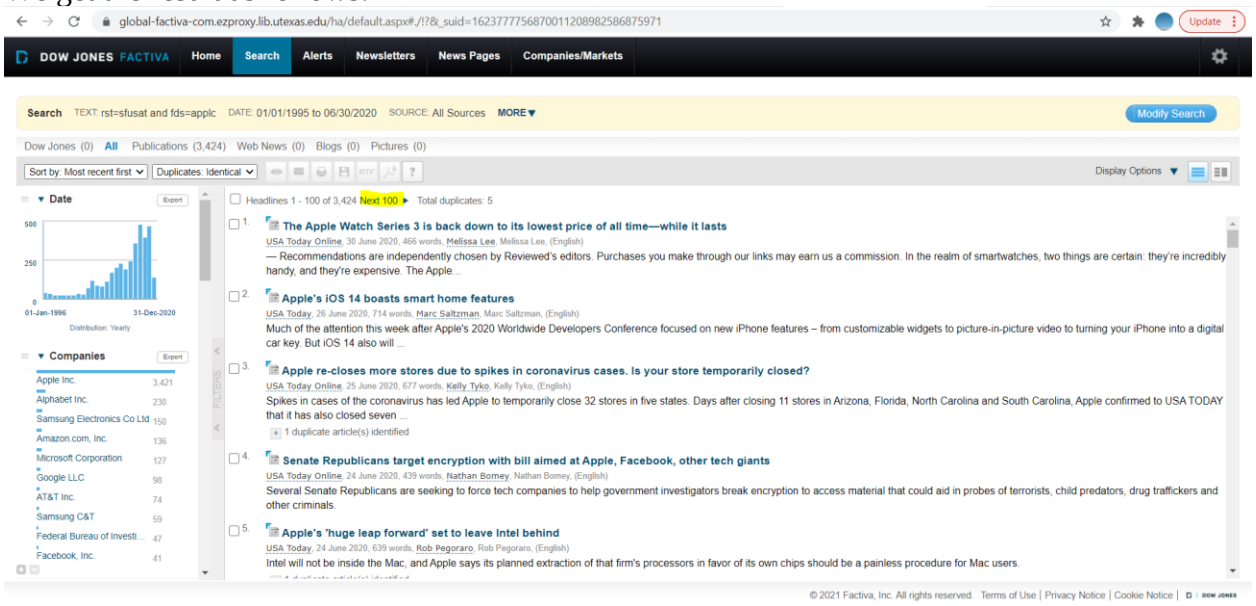
EXAMPLE 1: to search for the frequency that Apple Inc. is mentioned in *USA Today (All sources)*, i.e., newspaper_code = sfusat, we set the search as follows:

- Enter Date range: 01/01/1995 – 06/30/2020
- Search syntax: ***rst=sfusat and fds=applc***
- Set languages: all languages
- Set Duplicates: Identical

then click "Search".



We get the result as follows:



Then we can copy paste the page to text file and have the code run to count the frequency. The differences compared to sector frequency are:

(1) The information the code needs to search for is company (fds), not sector (in).

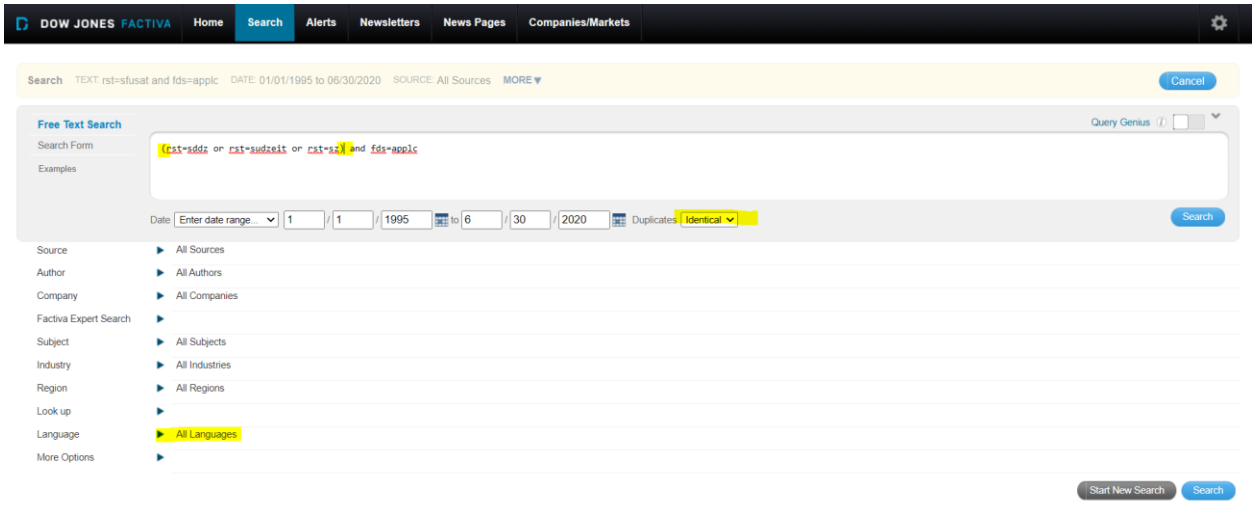
(2) Each RA will be in charge of several companies in the file 'STEP4_Factiva_company_list.xlsx". It means that the list of companies is not the same for different RAs (not like sector frequency: we have same list of sectors for everyone, and it is only 139 sectors).

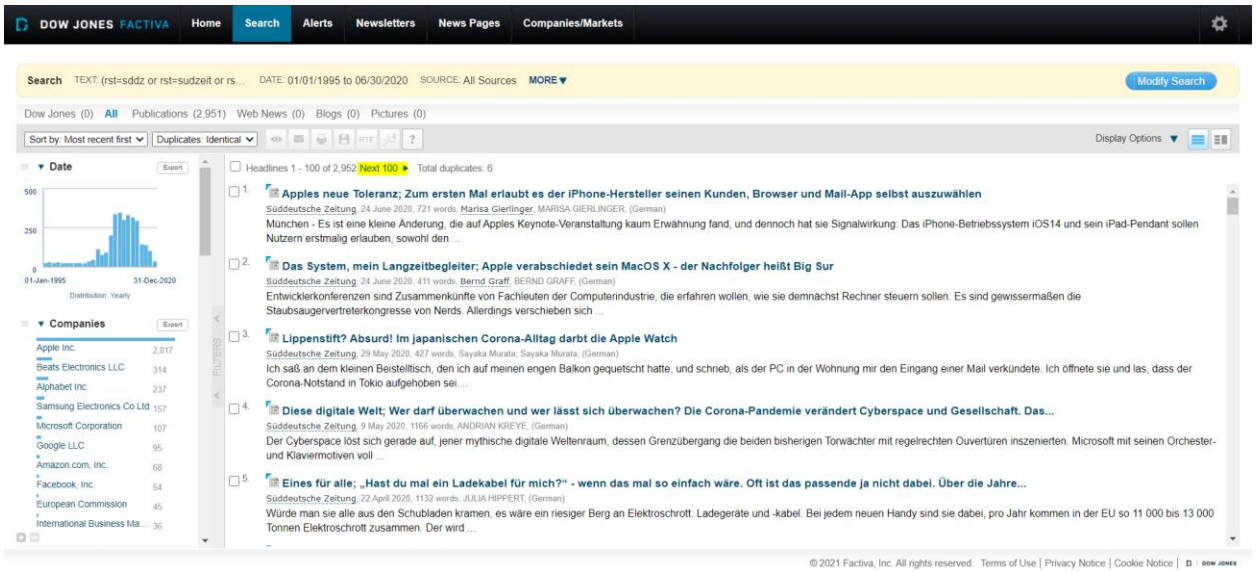
(3) We need to find newspaper_code in text file and match it in excel file this time. For example, the code needs to be able to identify "rst=sfusat" → match to "sfusat" of that company, or identify "(rst=lf or rst=figaro or rst=figbour or rst=figetu or rst=figmag or rst=figpre)" → match to "sffigarop" of that company. The Python code written by Alex will accommodate this part.

EXAMPLE 2: to search for the frequency that Apple Inc. is mentioned in Suddeutsche Zeitung (All sources) (language: German/English), we set the search as follows:

- Enter Date range: 01/01/1995 – 06/30/2020
- Search syntax: ***(rst=sddz or rst=sudzeit or rst=sz) and fds=applc***
- Set languages: all languages
- Set Duplicates: Identical

then click "Search".

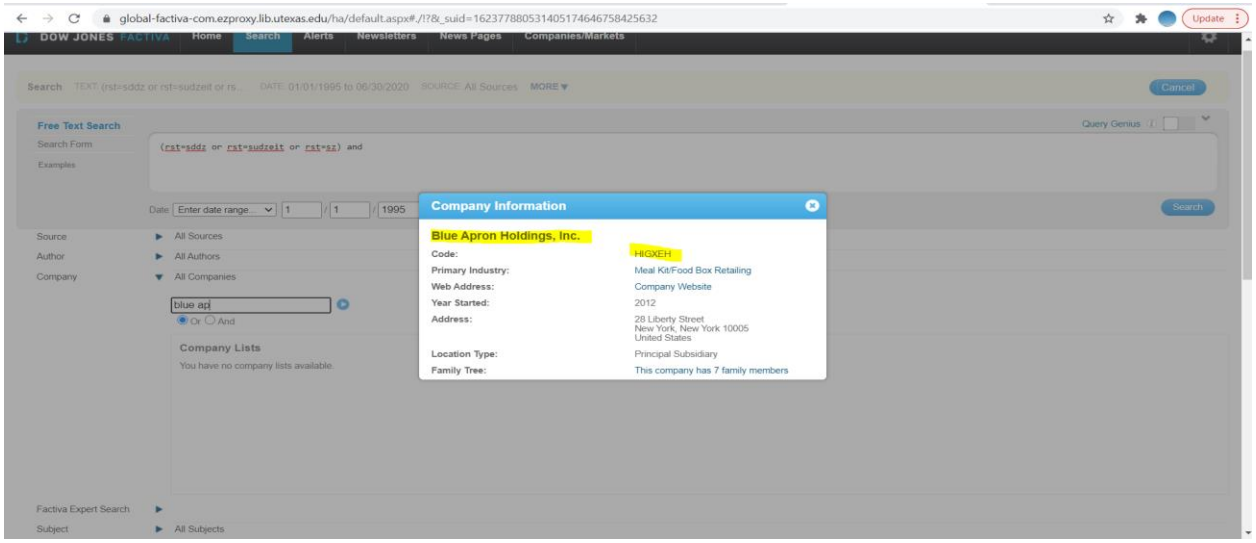


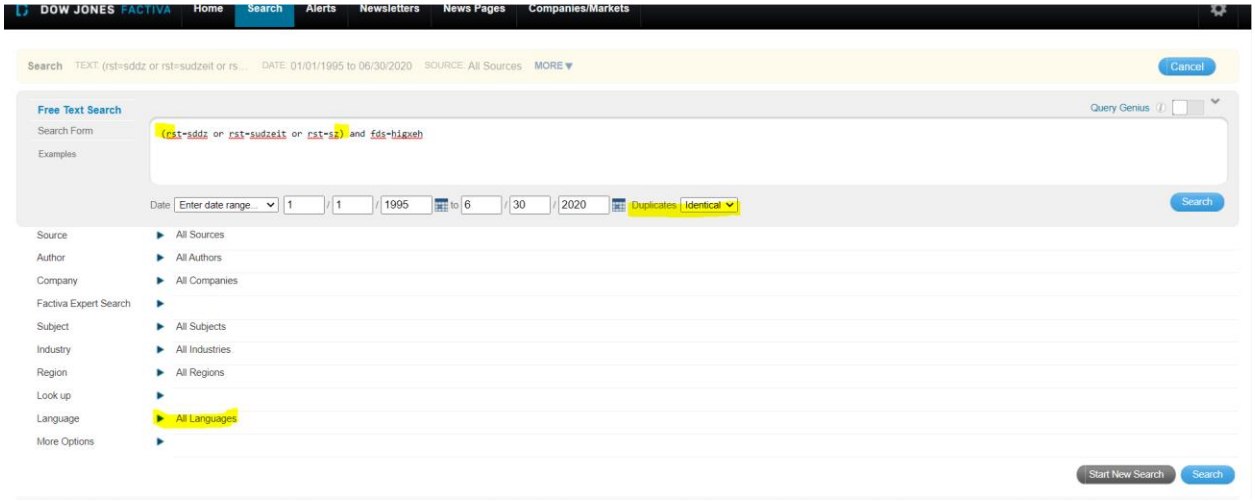


EXAMPLE 3: to search for the frequency that Blue Apron Holdings, Inc. is mentioned in Suddeutsche Zeitung (All sources) (language: German/English), we set the search as follows:

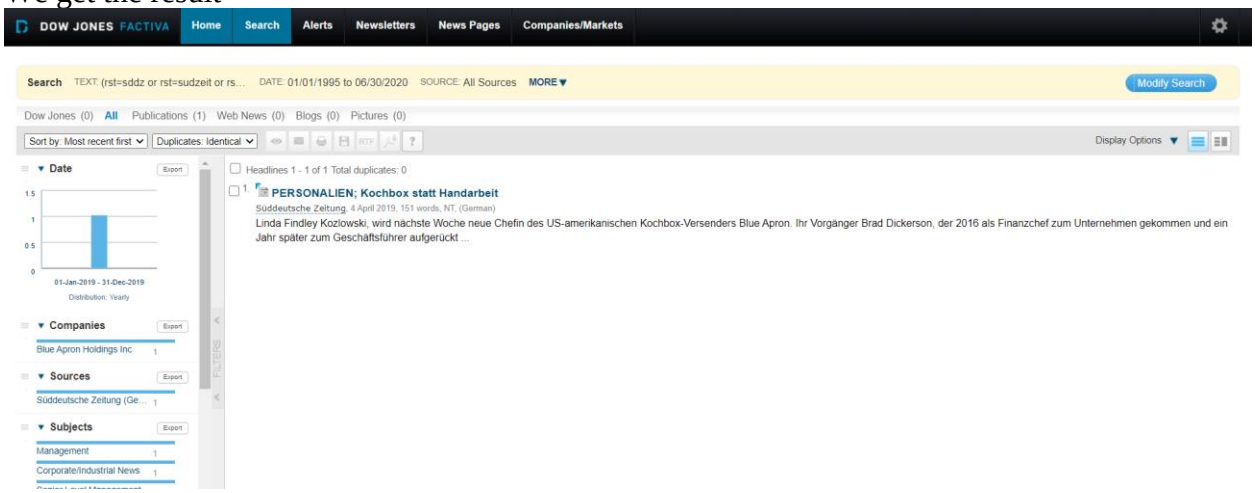
- Enter Date range: 01/01/1995 – 06/30/2020
- Search syntax: ***(rst=sddz or rst=sudzeit or rst=sz) and fds=higxeh***
- Set languages: all languages
- Set Duplicates: Identical

then click “Search”.





We get the result



⇨ Final product of this work is a file named "STEP4_Company_frequency_No<NUMBER 1>_<NUMBER 2>_YOURNAME.xlsx"
For example: **STEP4_Company_frequency_No1_20_Ha.xlsx**

IV. UPLOAD RESULTS AND WEEKLY REPORT

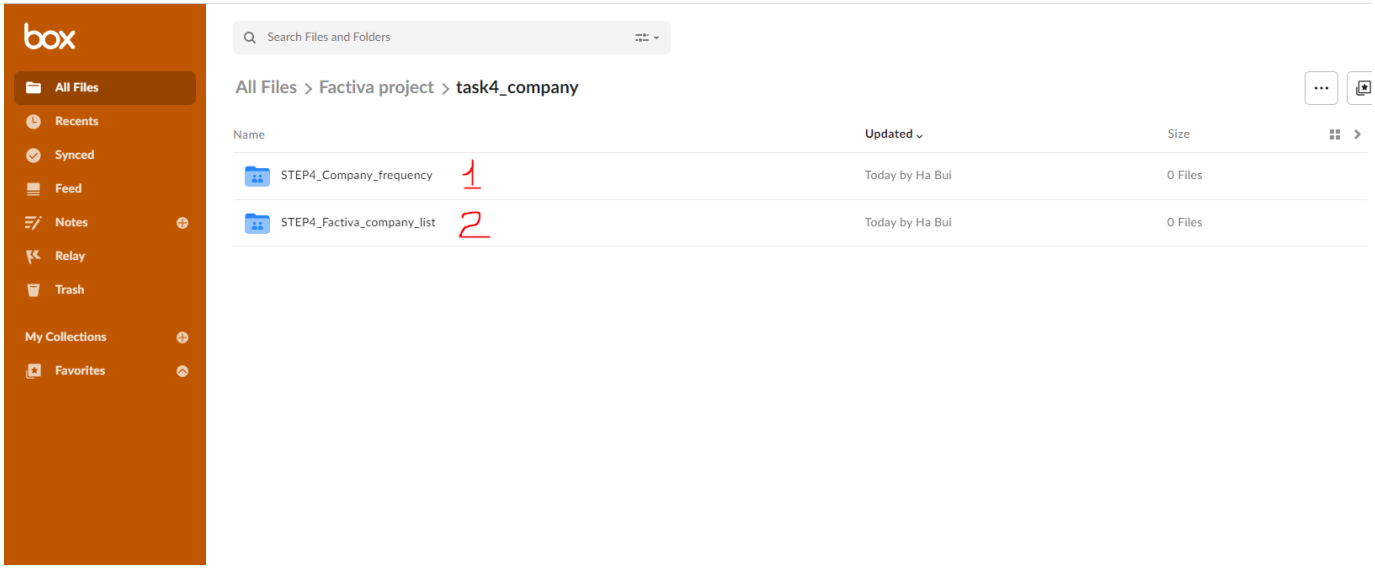
- Every Sunday, you will upload your files to Box. There are 2 files you created, and they should be uploaded to two folders 1 and 2 below:

1. Folder "STEP4_Company_frequency": upload File "STEP4_Company_frequency_No<NUMBER 1>_<NUMBER 2>_YOURNAME.xlsx"
For example: **STEP4_Company_frequency_No1_20_Ha.xlsx**

2. Folder "STEP4_Factiva_company_list": upload File "STEP4_Factiva_company_list_No<NUMBER 1>_<NUMBER 2>_sector_YOURNAME.xlsx"
For example: "**STEP4_Factiva_company_list_No1_20_sector_Ha.xlsx**":

- Together with uploading file to Box, please send me a quick email to factivaproject@gmail.com letting me know:
  - How many companies you finish in file "**STEP4_Factiva_company_list……"**
  - How many companies you finish in file "**STEP4_Company_frequency……..".** If you are in the middle of a company, please let me know as well.

For example: This week I got information for 20 companies in file "STEP4_Factiva_company_list……" and finished 10 companies in 11 newspapers (Company ID 101-110). I finished sfusat, sfwsj, sfnyt for company ID 111.



V. DUPLICATES FACTIVA_COMPANY_NAME1

Please notice that there are 280 companies that have duplicates (they are the first 280 observations in file "STEP4_Factiva_company_list.xlsx"). They can be true duplicates or just 2 companies having the same names but different factiva_company_code. For example: there are actually two **acma ltd** companies in 2 different countries with different factiva_company code

	factiva_company_name1	factiva_company_code	company_country
1	aalberts nv		
2	aalberts nv	ALBTS	Netherlands
3	aberdeen greater china fund inc		
4	aberdeen greater china fund inc	GRCHIN	USA
5	acma ltd	ACMAEI	Singapore
6	acma ltd	ACMGOL	UK
7	adt inc	ADTSSI	USA
8	adt inc		
9	advanta corp		
10	advanta corp	TSOF	USA
11	allied healthcare products inc		
12	allied healthcare products inc	ALHLTH	USA
13	ambu as		
14	ambu as	AMBUDK	Denmark
15	american medical security group inc		
16	american medical security group inc	UWISC	USA
17	american physicians service group inc		
18	american physicians service group inc	APSC	USA
19	asat holdings ltd		
20	asat holdings ltd	ASATIN	Hong Kong
21	asia holdings co ltd		
22	asia holdings co ltd	ASCEMM	South Korea
23	asia pacific fund inc	APACFI	USA
24	asia pacific fund inc		
25	asseco poland sa		
26	asseco poland sa	SFTK	Poland
27	atlantia spa		
28	atlantia spa	AUTSD	Italy
29	azure midstream partners lp		
30	azure midstream partners lp	MLMIDZ	USA
31	bank bph sa		
32	bank bph sa	BPRZH	Poland
33	bell aliant regional communications inc		

Solution:

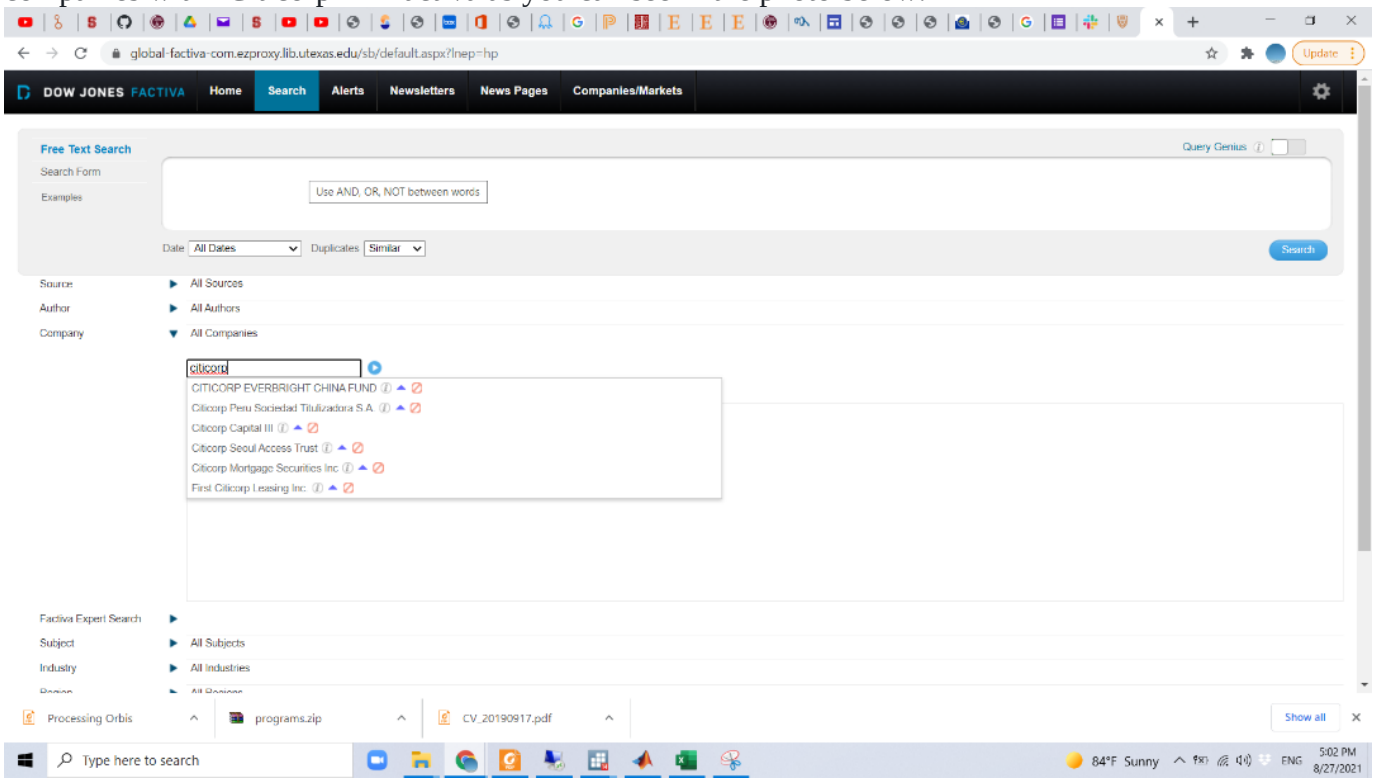
- If it is the case like "**acma ltd"** then please go ahead and do two companies like two separate companies. Since they have different factiva_company_code, the search syntax will not confuse those two companies.
- If it is actually duplicate: you only need to do for 1 observation (i.e. (1) only fill in information for 1 company in file "STEP4_Factiva_company_list_No<NUMBER 1>_<NUMBER

2>_sector_YOURNAME.xlsx", (2) only search for frequency of that one company in file "STEP4_Company_frequency_No<NUMBER 1>_<NUMBER 2>_YOURNAME.xlsx". I will read from your file "STEP4_Factiva_company_list_No<NUMBER 1>_<NUMBER 2>_sector_YOURNAME.xlsx" to know that they are true duplicate.

If you work on a company outside those 280 companies and find that there are 2 or more companies with the same name in Factiva (but they are actually different, i.e., have different factiva_company_code), please do the following things:
(1) add that new company to your file "STEP4_Factiva_company_list_No<NUMBER 1>_<NUMBER 2>_sector_YOURNAME.xlsx", obtain information like normal,
(2) search for frequency of both companies like normal and put into file "STEP4_Company_frequency_No<NUMBER 1>_<NUMBER 2>_YOURNAME.xlsx",
(3) please let me know in your weekly progress email that you found a new company.

VI. SPECIAL CASE OF COMPANY NAME

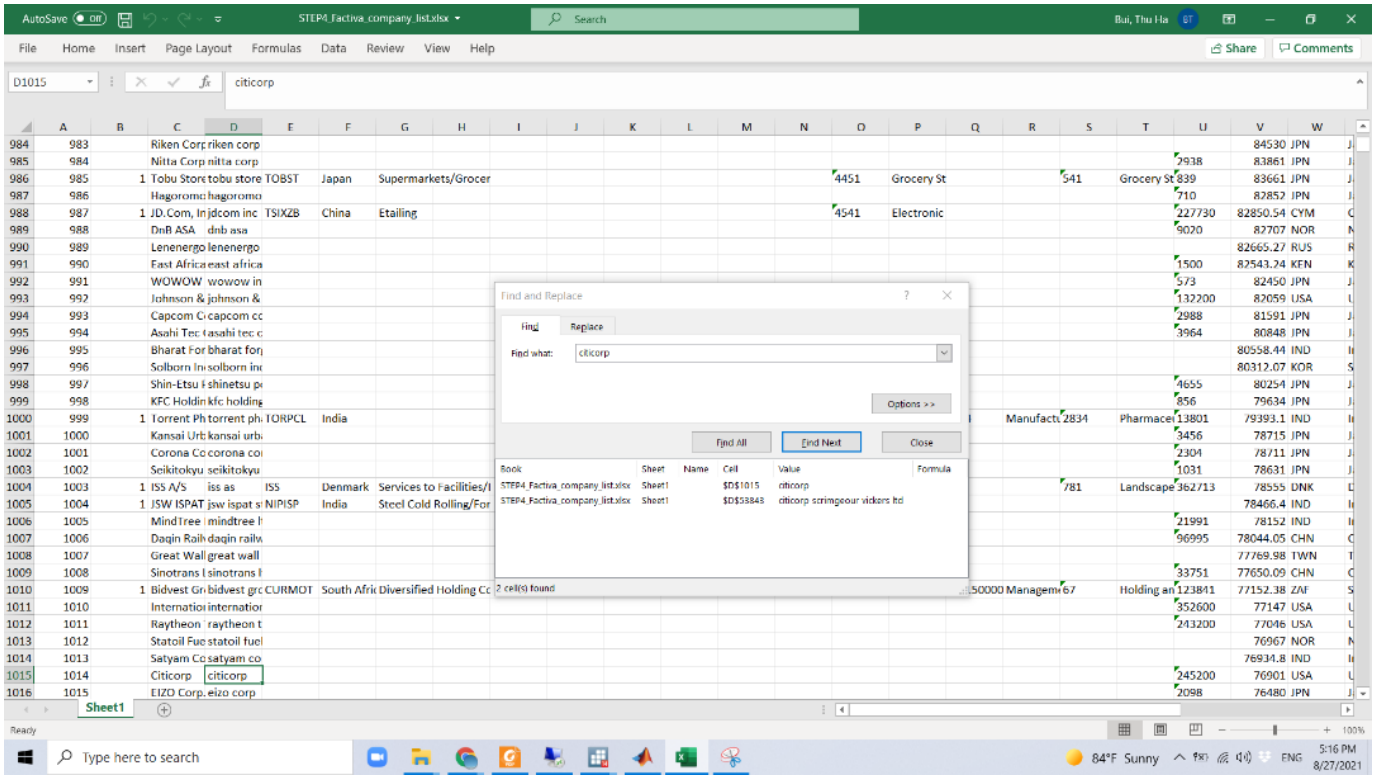
It is the case when you work on 1 company name, but actually find multiple company names in Factiva. For example, we have a firm named "Citicorp" in file Factiva_company_list. However, there are several companies with "Citicorp" in Factiva as you can see in the photo below.



What you will do are:

Step 1: check the big file STEP4_Factiva_company_list.xlsx" if those companies exit in the list.

For example, in this case, we only see 2 companies in the list that have "citicorp" in their names. The reason is because the list you are using is data we collected in previous step using our method, it is not an exhaustive list of companies provided by Factiva. Our target is to get all companies, but it seems our method in previous step could not capture all companies in Factiva. So it is great that you are doing Task 4 and we could capture those companies now.



Step 2: Now you have identified that there are companies in Factiva but not in the list, please do as follows:
- For your file "Factiva company list (your file with 50 companies)": add new rows for other citicorp companies, add information like normal. In order to not mess up the ID, please put 1 company name in the place of that citicorp and add the others at the end. For those added company names at the end, please don't add ID, I will handle ID for those companies later.

- For your file "Company frequency": please collect frequency for all those citicorp you can find.

Step 3: In your weekly progress email, please notify me of these cases and if you observe any interesting information about these firms, please let me know as well (for example, those companies mostly went inactive).

VII. MISSING FCODEINCODE AND COMPANY INFORMATION

- Chris has notified me two issues: (1) missing fcodeincode information after doing the merge (in python/R) to get file "STEP4_Factiva_company_list_No..._..._sector_....xlsx", and (2) missing company information in file "STEP4_Factiva_company_list_No..._..._sector_....xlsx".

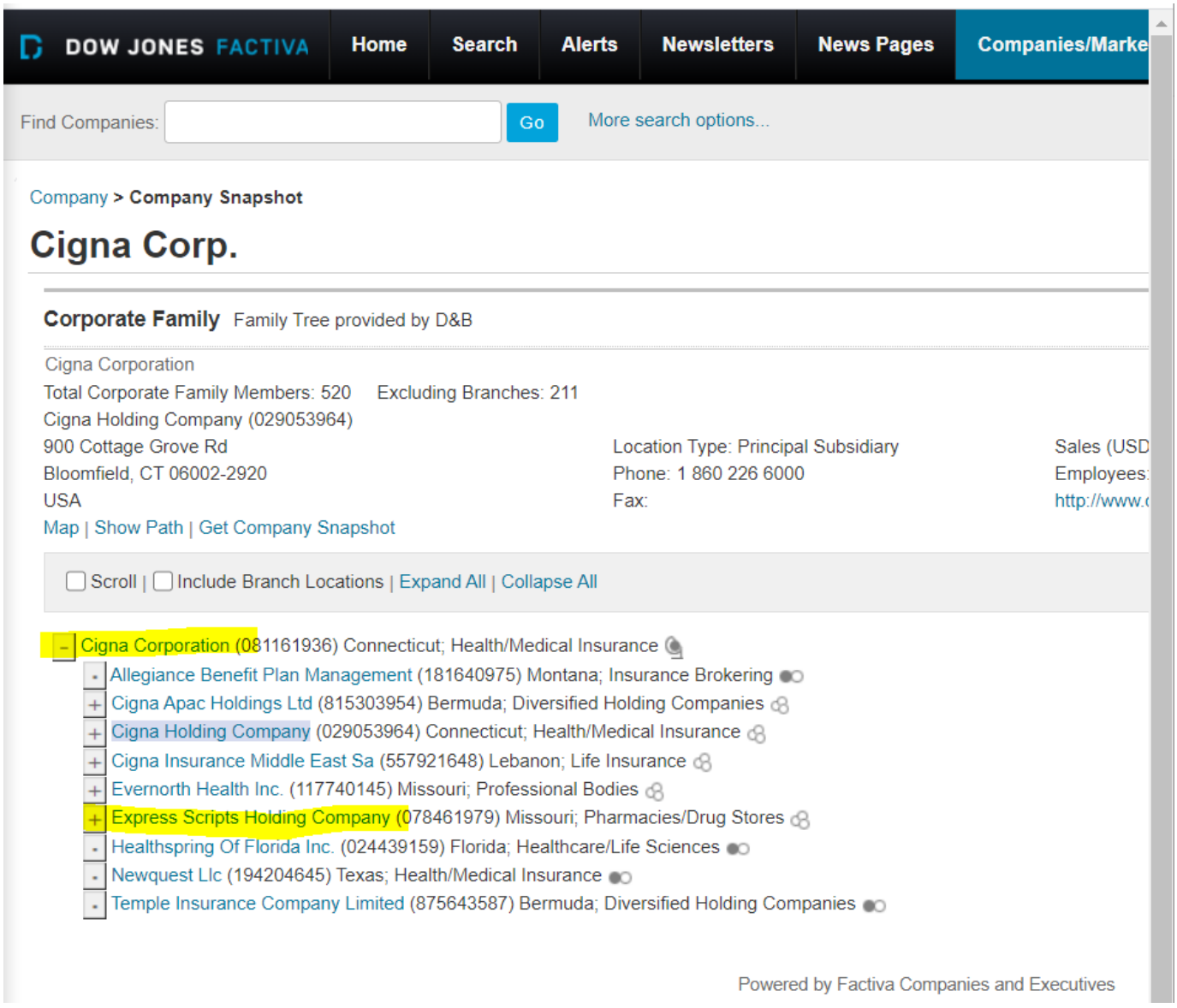
I have recorded a video to provide solutions for these two issues. The recording is accessible via this link https://utexas.zoom.us/rec/share/x0jPKmzLNc4skN8yXkB02szLB3uekCWK4_E47u_miC8RoC4nIPXIv_WF2E914Ohb.VrZawwmfrhlandPT. I also put the recording into Box, folder 'Factiva guide".

- Continuing my previous email on missing company information. Chris proposed some suggestions on missing information for column "location_type", "global_ultimate", "domestic_ultimate", "immediate_parent" and I think that they are great suggestion so I would like to share with you all.

Based on Chris's observation, there are four distinct cases to deal with:

1) The company has been acquired and still exists as a subsidiary of another company (Dangote Flour and Express Scripts) -- Express Scripts is the one I talked about in the recording.
2) The company has gone bankrupt and no longer exists (Takefuji Corp)
3) The company has merged on equal terms with another company and changed its name but is not a subsidiary per se (Shell)
4) The company has been acquired and its operation incorporated into another company and no longer exists even as a subsidiary (United Bank of India)

For case (1): you can put:
- location_type: principal subsidiary
- "global_ultimate", "domestic_ultimate", "immediate_parent": please list its new parent company. For example, for Express Scripts, you can list "Cigna Corporation"



For cases 2, 3, 4: you can put:
- location_type: headquarters
- "global_ultimate", "domestic_ultimate", "immediate_parent": list the company as its own parent. For example, I would put "Takefuji Corp" for those 3 columns for Takefuji Corp.