

# User Manual

## sec-10-k-scraper

Creators: Ed Carl, Chloe Dorward, Seth Markarian, Sena Yevenyo

An up-to-date version of this document can always be found [on Google Docs](#).

A video demonstration of the app can be found [here](#).

# Summary

## Purpose of Application

This app is a local desktop application for Windows, MacOS, and Linux that obtains companies' 10-K, 10-Q, and 10-F SEC filings, downloads the filings as HTML files, extracts the text of key sections of those documents, analyzes the text for named entities, and gathers that information in a Microsoft Excel spreadsheet. Users can either search for filings based on company names/ CIK numbers, filing type, and date range, or upload a file including similar parameters for batch uploading.

Using a desktop application rather than a web service minimizes the technical and financial requirements for FracTracker to operate it. The application is entirely self-contained on users' computers and there will be no need to spend time on configuring it or money on hosting a server. A user only needs to have an internet connection in order to run the application.

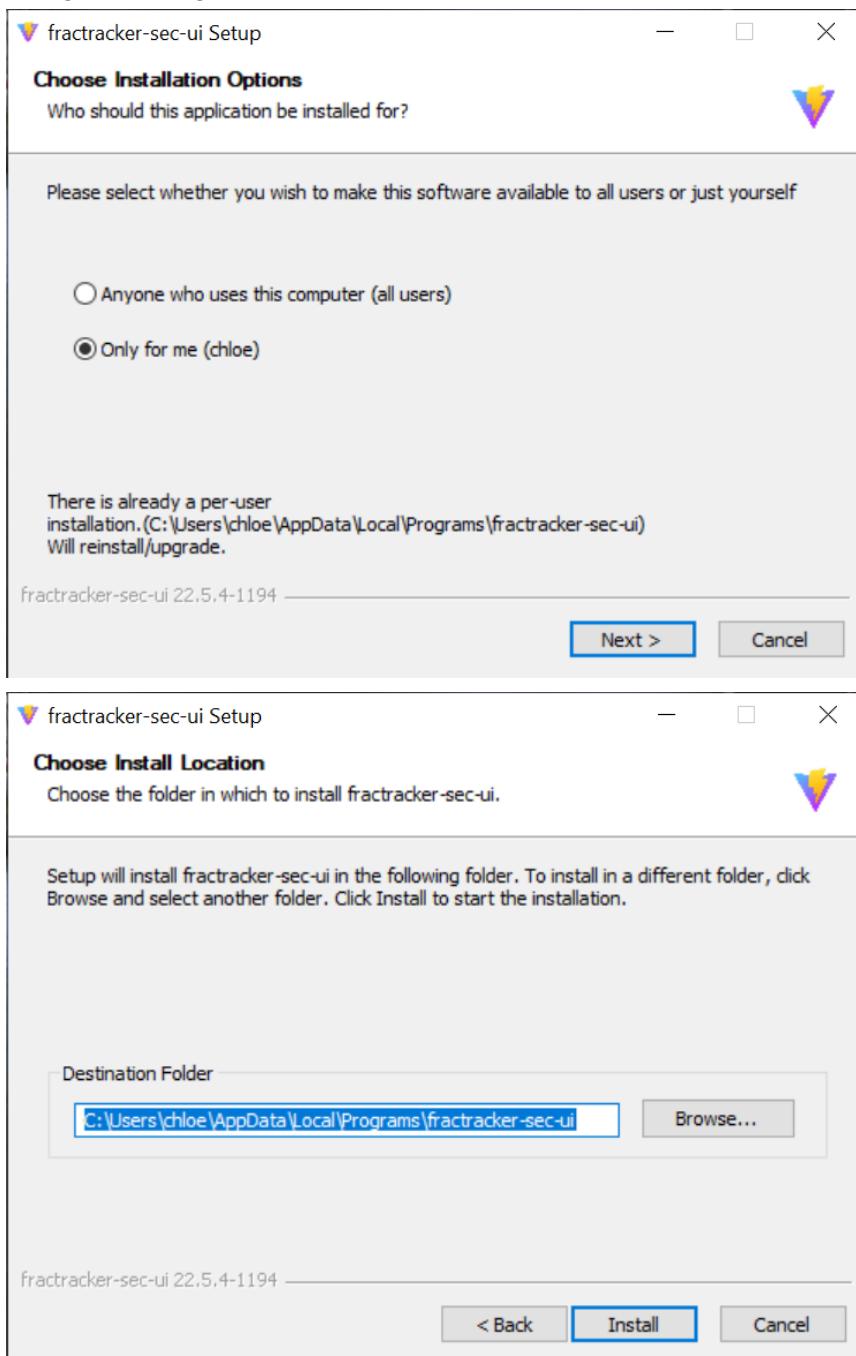
# Using the App

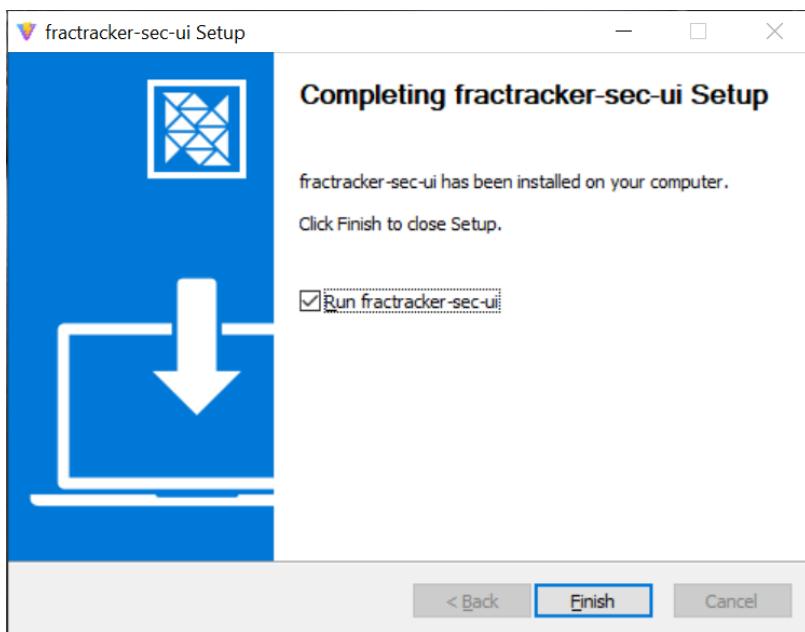
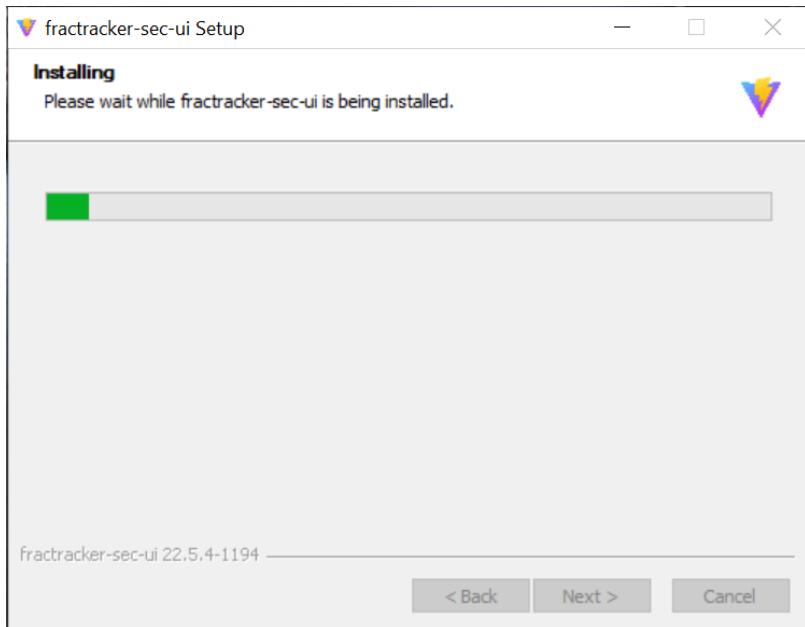
## Installing & Running the App

Releases of the application installer can be found under the releases page of the project's github repository. [github.com/edwardmcarl/sec-10-k-scraper/releases](https://github.com/edwardmcarl/sec-10-k-scraper/releases)

### Windows

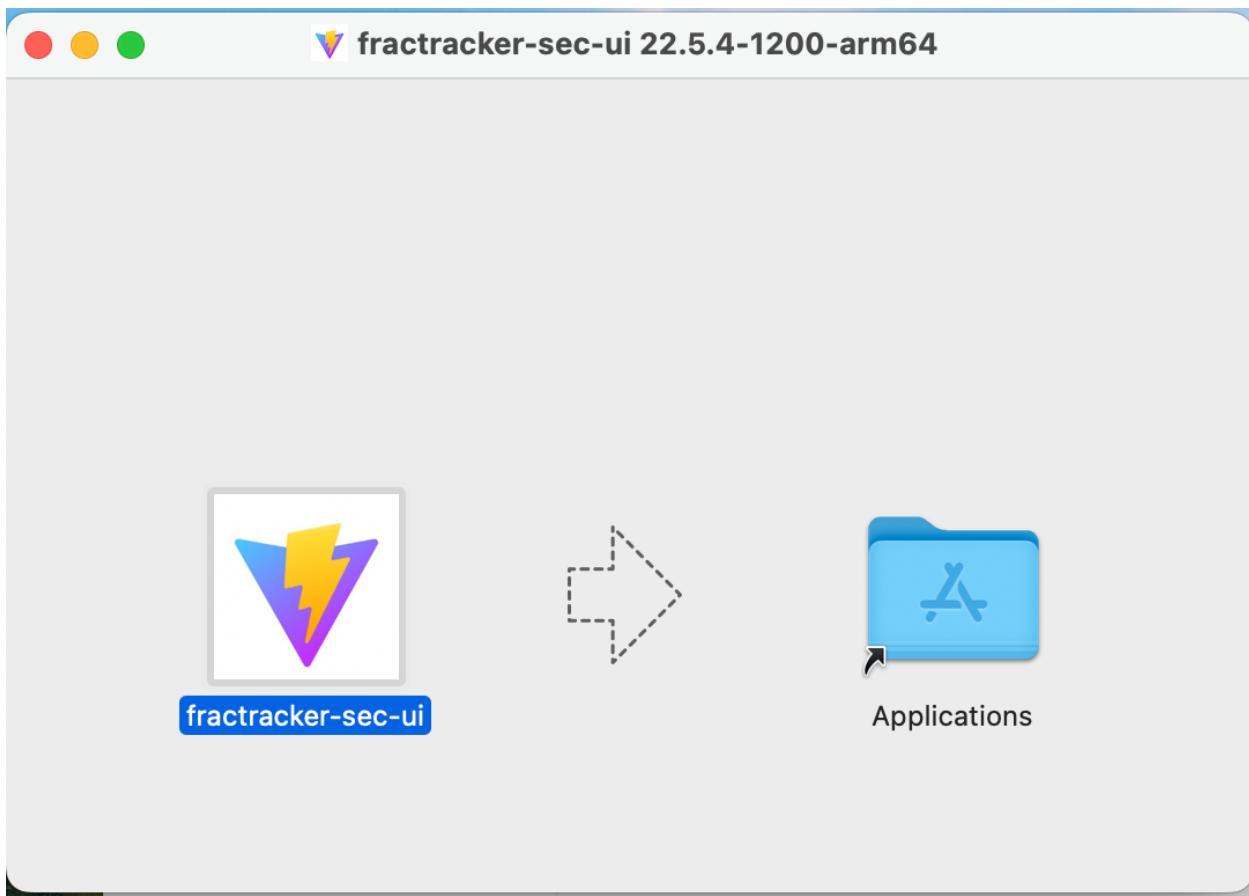
Find the installer in your file explorer (fractracker-sec-ui Setup 22.5.4-1194.exe), and run it. Navigate through the prompts to install.





An icon will appear on your desktop.

Mac



Simply mount the installer .dmg and drag the app into the Applications folder.

Linux

Just double-click the installer ApplImage.

## Step 1: Selecting Filings

### Option 1: Searching

#### Searching

The first option is to select an entity, and search for filings based on a date range.



## SEC EDGAR 10-K Information Access

ford

CIK0000038264 | Forward Industries, Inc. (FORD)

CIK0000037996 | FORD MOTOR CO (F, F-PC, F-PB)

CIK0001586097 | Ford Tim

CIK0001176401 | FORD JOE T

CIK0001216553 | FORD BETH

CIK0001511094 | Ford Adam

CIK0001390145 | Ford Jill

CIK0001606824 | Ford Mark

CIK0001392324 | Ford Eric

CIK0001143133 | FORD MELBA

Click on the search bar, and begin to input the appropriate entity name or CIK number. A list of predicted entities will appear in a list below the bar.



## SEC EDGAR 10-K Information Access

FORD MOTOR CO (F, F-PC, F-PB)

Start Date: 5/2/2021

End Date: 5/2/2022

Form Type:

Read from file (.txt):

No file chosen

Entity Name	CIK Number	Form Type	Filing Date	Link to Document	Extract Info?
FORD MOTOR CO	CIK0000037996	10-K	2022-02-04	<a href="https://sec.gov/Archives/edgar/data/37996/000003799622000013/f-20211231.htm">https://sec.gov/Archives/edgar/data/37996/000003799622000013/f-20211231.htm</a>	<input type="button" value="Add to Queue"/>

Select an entity by clicking on it (for example, FORD MOTOR CO). This will populate the results table with the default search parameters - 10-K filings from the past year.

Start Date: 5/2/2021   End Date: 5/2/2022

Read from

Entity Name	CIK Number	Form Type	Filing Date	Link to Document
FORD MOTOR CO	CIK0000037996	10-K	2022-02-04	<a href="https://sec.gov/Archives/edgar/data/37996/000003799622000013/f-20211231.htm">https://sec.gov/Archives/edgar/data/37996/000003799622000013/f-20211231.htm</a>

« < May 2021 > »

MON	TUE	WED	THU	FRI	SAT	SUN
26	27	28	29	30	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31	1	2	3	4	5	6

You can adjust your date range by clicking on the date selector, and changing the date accordingly.

Start Date: 5/2/2022 End Date: 5/2/2022

Read from

Entity Name	CIK Number	Form Type	Filing Date

You can also type in the date rather than using the provided selector.

Form Type:

10-K

10-Q

20-F

You can change the selected form type by clicking on the gray dropdown on the right hand side, and clicking on the appropriate form.

Entity Name	CIK Number	Form Type	Filing Date	Link to Document	Extract Info?
FORD MOTOR CO	CIK0000037996	10-K	2022-02-04	<a href="https://sec.gov/Archives/edgar/data/37996/000003799622000013/f-20211231.htm">https://sec.gov/Archives/edgar/data/37996/000003799622000013/f-20211231.htm</a>	<input type="button" value="Add to Queue"/>
FORD MOTOR CO	CIK0000037996	10-K	2021-02-05	<a href="https://sec.gov/Archives/edgar/data/37996/000003799621000012/f-20201231.htm">https://sec.gov/Archives/edgar/data/37996/000003799621000012/f-20201231.htm</a>	<input type="button" value="Add to Queue"/>
FORD MOTOR CO	CIK0000037996	10-K	2020-02-05	<a href="https://sec.gov/Archives/edgar/data/37996/000003799620000010/f1231201910-k.htm">https://sec.gov/Archives/edgar/data/37996/000003799620000010/f1231201910-k.htm</a>	<input type="button" value="Add to Queue"/>

Clicking on the blue Search button will repopulate the results table based on your selected parameters. Changing the date, form type, or search components also repopulates the results.

## Selecting Filings

Entity Name	CIK Number	Form Type	Filing Date	Link to Document	Extract Info?
FORD MOTOR CO	CIK0000037996	10-K	2022-02-04	<a href="https://sec.gov/Archives/edgar/data/37996/000003799622000013/f-20211231.htm">https://sec.gov/Archives/edgar/data/37996/000003799622000013/f-20211231.htm</a>	<input type="button" value="Add to Queue"/>
FORD MOTOR CO	CIK0000037996	10-K	2021-02-05	<a href="https://sec.gov/Archives/edgar/data/37996/000003799621000012/f-20201231.htm">https://sec.gov/Archives/edgar/data/37996/000003799621000012/f-20201231.htm</a>	<input type="button" value="Add to Queue"/>
FORD MOTOR CO	CIK0000037996	10-K	2020-02-05	<a href="https://sec.gov/Archives/edgar/data/37996/000003799620000010/f1231201910-k.htm">https://sec.gov/Archives/edgar/data/37996/000003799620000010/f1231201910-k.htm</a>	<input type="button" value="Add to Queue"/>

Individual filings from different years can be selected by clicking on the Add to Queue button.

Entity Name	CIK Number	Form Type	Filing Date	Link to Document	Extract Info?
FORD MOTOR CO	CIK0000037996	10-K	2022-02-04	<a href="https://sec.gov/Archives/edgar/data/37996/000003799622000013/f-20211231.htm">https://sec.gov/Archives/edgar/data/37996/000003799622000013/f-20211231.htm</a>	<input type="button" value="Remove from Queue"/>
FORD MOTOR CO	CIK0000037996	10-K	2021-02-05	<a href="https://sec.gov/Archives/edgar/data/37996/000003799621000012/f-20201231.htm">https://sec.gov/Archives/edgar/data/37996/000003799621000012/f-20201231.htm</a>	<input type="button" value="Remove from Queue"/>
FORD MOTOR CO	CIK0000037996	10-K	2020-02-05	<a href="https://sec.gov/Archives/edgar/data/37996/000003799620000010/f1231201910-k.htm">https://sec.gov/Archives/edgar/data/37996/000003799620000010/f1231201910-k.htm</a>	<input type="button" value="Remove from Queue"/>

Individual filings can be removed by clicking on the Remove from Queue button.  
Repeat the process outlined above to add filings from different entities to the same Queue.

### Option 2: Uploading

The second option is to upload a text file with different CIK numbers and appropriate date ranges. This option will add every filing within a CIK's date range to the Queue.

#### File Format:

The file must be a .txt file. Your computer will have a program to create text files, for example, Notepad,TextEdit, etc.

Each line represents a new entity, and should contain the CIK number, the start of the date range, and the end of the date range:

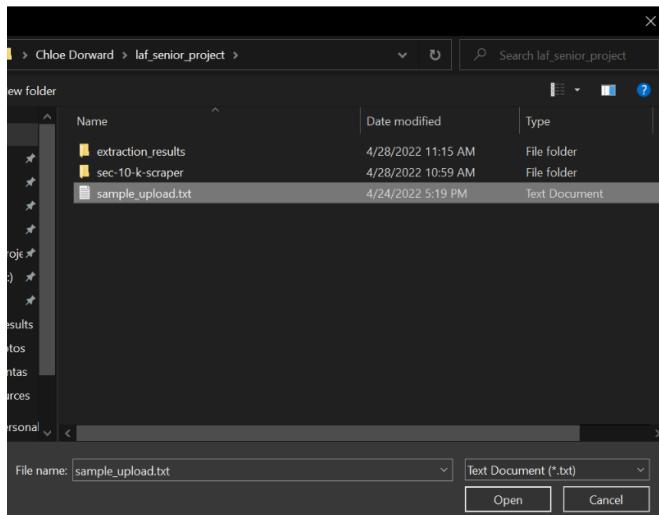
CIK00000XXXXX YYYY-MM-DD YYYY-MM-DD

### Uploading the File

Read from file (.txt): [\(i\)](#)

No file chosen

Click on the Choose File button.



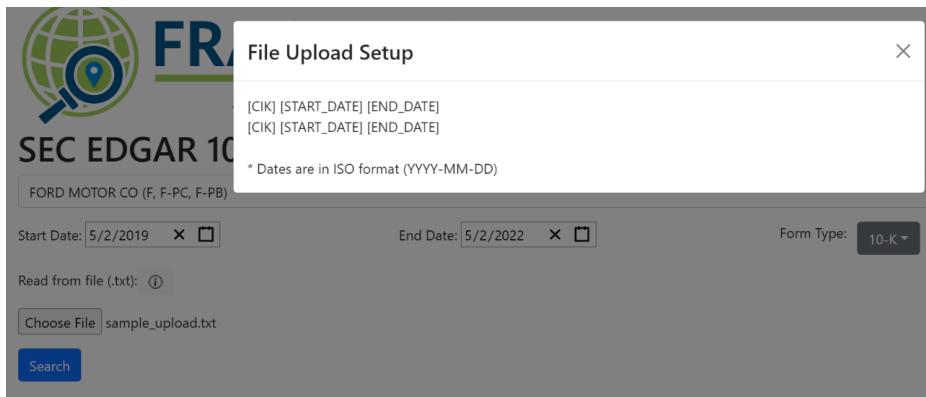
Navigate through your file system to select your file.

Read from file (.txt): [\(i\)](#)

sample\_upload.txt

The uploaded file name will appear next to the Choose File button.

This option can be used in combination with Option 1. Any results that have already been selected through the File Upload option will appear as such in the search results table.



File Upload Setup

[CIK] [START\_DATE] [END\_DATE]  
[CIK] [START\_DATE] [END\_DATE]

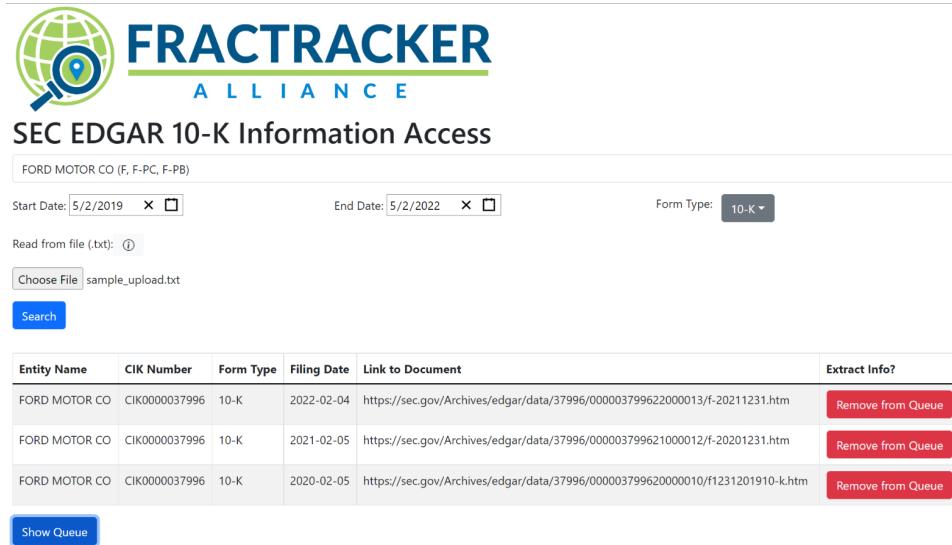
\* Dates are in ISO format (YYYY-MM-DD)

Start Date: 5/2/2019 End Date: 5/2/2022 Form Type: 10-K

Read from file (.txt):  sample\_upload.txt

If you forget the format for file uploading, it can be found by clicking on the i button.

## Step 2: Viewing Queue



FRACTRACKER  
ALLIANCE

### SEC EDGAR 10-K Information Access

FORD MOTOR CO (F, F-PC, F-PB)

Start Date: 5/2/2019 End Date: 5/2/2022 Form Type: 10-K

Read from file (.txt):  sample\_upload.txt

Entity Name	CIK Number	Form Type	Filing Date	Link to Document	Extract Info?
FORD MOTOR CO	CIK0000037996	10-K	2022-02-04	<a href="https://sec.gov/Archives/edgar/data/37996/000003799622000013/f-20211231.htm">https://sec.gov/Archives/edgar/data/37996/000003799622000013/f-20211231.htm</a>	<input type="button" value="Remove from Queue"/>
FORD MOTOR CO	CIK0000037996	10-K	2021-02-05	<a href="https://sec.gov/Archives/edgar/data/37996/000003799621000012/f-20201231.htm">https://sec.gov/Archives/edgar/data/37996/000003799621000012/f-20201231.htm</a>	<input type="button" value="Remove from Queue"/>
FORD MOTOR CO	CIK0000037996	10-K	2020-02-05	<a href="https://sec.gov/Archives/edgar/data/37996/000003799620000010/f1231201910-k.htm">https://sec.gov/Archives/edgar/data/37996/000003799620000010/f1231201910-k.htm</a>	<input type="button" value="Remove from Queue"/>

View the Queue by clicking on the blue Show Queue button on the bottom of the page.

The screenshot shows the Fractracker Alliance SEC EDGAR 10-K Information Access interface. On the left, there's a search form with fields for Start Date (5/2/2019) and End Date (5/2/2022). Below that is a table of search results for FORD MOTOR CO, showing columns for Entity Name, CIK Number, Form Type, Filing Date, and Link to Document. On the right, a separate window titled 'Queue' lists filings with columns for Entity Name, CIK Number, Form Type, Filing Date, and a red 'Remove From Queue' button.

This will slide open the Queue tab.

Entity Name	CIK Number	Form Type	Filing Date	
FORD MOTOR CO	CIK0000037996	10-K	2022-02-04	<span style="background-color: red; color: white; padding: 2px;">Remove From Queue</span>

From here, you can remove filings from the queue.

## Step 3: Downloading Files

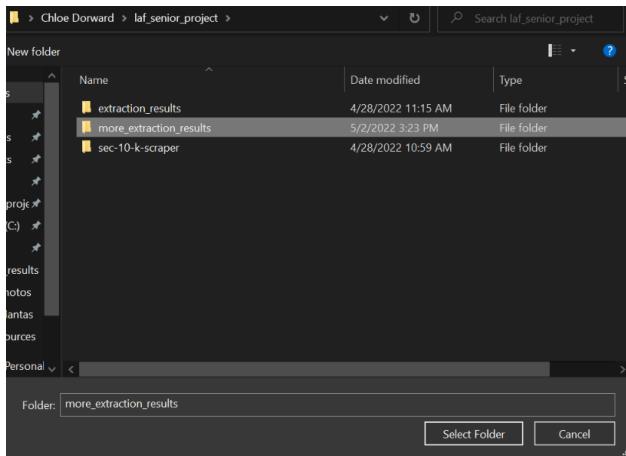
Queue

Choose Path for Download  
C:\Users\chloe\Desktop

Apply Named Entity Recognition to Queue

Extract & Download

At the top of the Queue tab, click the Choose Path button. This allows you to select a place to save your extracted files. The default path is your Desktop folder.



Navigate through the file dialog, and select a folder to save the files to.

#### Queue

Choose Path for Download  
C:\Users\chloe\laf\_senior\_project\more\_extraction\_results  
 Apply Named Entity Recognition to Queue  
**Extract & Download**

Once a location is selected, the name of the file path will appear below the Choose File button.

#### Queue

Choose Path for Download  
C:\Users\chloe\laf\_senior\_project\more\_extraction\_results  
 Apply Named Entity Recognition to Queue  
**Extract & Download**

Next, you have the option to either apply Named Entity Recognition during the file extraction process, or not. You can select the checkbox to do this.

#### Queue

Choose Path for Download  
C:\Users\chloe\laf\_senior\_project\more\_extraction\_results  
 Apply Named Entity Recognition to Queue  
**Extract & Download**

Finally, press the Extract & Download button.

The screenshot shows the Fractracker Alliance SEC EDGAR 10-K Information Access interface. On the left, there's a search bar for 'FORD MOTOR CO (F, F-PC, F-PB)' and date filters ('Start Date: 5/2/2019' and 'End Date: 5/2/2022'). Below these are fields for reading from a file (.txt) and a 'Search' button. A table lists entities with columns: Entity Name, CIK Number, Form Type, Filing Date, and Link to Document. The table contains entries for Ford Motor Co and Hershey Co. On the right, a 'Queue' window is open, showing a table of extracted results with columns: Entity Name, CIK Number, Form Type, Filing Date, and a 'Remove From Queue' button. The table includes rows for Ford Motor Co, Hershey Co, and Continental Resources, Inc.

While the system is extracting the files, it will indicate that it is loading. The buttons next to the selected filings will gray out, meaning that they are processing. You cannot deselect an entity once the extraction process has started.

Once the extraction process is complete, the loading spinner will disappear, and the selected filings will no longer be grayed out.

## Step 4: Finding/Using Your Files

### File Structure

Files will be downloaded to the folder that you selected in Step 3.

The screenshot shows a file explorer window with the path: Chloe Dorward > laf\_senior\_project > more\_extraction\_results. The contents of this folder are listed in a table:

Name	Date modified	Type	Size
CONTINENTAL RESOURCES, INC	5/2/2022 3:24 PM	File folder	
FORD MOTOR CO	5/2/2022 3:24 PM	File folder	
HERSHEY CO	5/2/2022 3:24 PM	File folder	
summary.xlsx	5/2/2022 3:25 PM	Microsoft Excel Work...	95 KB

Within your selected folder, there will be a folder for each entity, and an excel spreadsheet.

The screenshot shows a file explorer window with the path: Chloe Dorward > laf\_senior\_project > more\_extraction\_results > CONTINENTAL RESOURCES, INC. The contents of this folder are listed in a table:

Name	Date modified	Type	Size
10-K_2020-02-26.htm	5/2/2022 3:24 PM	Chrome HTML Docu...	3,840 KB
10-K_2021-02-16.htm	5/2/2022 3:24 PM	Chrome HTML Docu...	3,591 KB

Each entity folder will contain htm files for each filing. The name of the file will be the filing type and the filing date.

## Excel File

A	B	C	D	E	F	G	H
1	Company Name	EIN	HQ Address of Incorporation	1. Business	NER 1. Business	Risk Factors	1A. Risk Factors
2	FORD MO 38054919({'street1': 'DE		See origin: []	See origin: ['Autoc			
3	FORD MO 38054919({'street1': 'DE		See origin: []	See origin: ['Autoc			
4	FORD MO 38054919({'street1': 'DE		See origin: []	See origin: ['Autoc			
5	HERSHEY (23069159({'street1': 'DE		>Item 1.B[]	>Item 1.A.[]['Amp			
6	HERSHEY (23069159({'street1': 'DE		>Item 1.B[]	>Item 1.A.[]['Amp			
7	CONTINENTAL 73076754({'street1': 'OK		See origin: []	See origin: ['Applic			
8	CONTINENTAL 73076754({'street1': 'OK		See origin: []	See origin: ['Audit			

Within the summary.xlsx excel file, each new filing will add a new row to the spreadsheet. If you choose to download your files to a folder which already has the summary.xlsx file, all additional filings will be appended to the end of the spreadsheet; our tool does not look through to check if there are duplicates.

The Columns in the spreadsheet are:

Company Name; EIN; HQ Address; State of Incorporation; 1. Business; NER 1. Business; 1A. Risk Factors; NER 1A. Risk Factors; 2. Properties; NER 2. Properties; 3. Legal Proceedings; NER 3. Legal Proceedings; 6. [Reserved]; NER 6. [Reserved]; 7. Financial Condition; NER 7. Financial Condition; 7A. Market Risk; NER 7A. Market Risk; 10. Corporate Governance; NER 10. Corporate Governance; 12. Security Ownership and Stockholder Matters; NER 12. Security Ownership and Stockholder Matters; 13. Certain Relationships; NER 13. Certain Relationships

Cells pertaining to entity information should always be filled in.

Cells pertaining to subsections of the filing will contain the text within that section if possible, or state: See original document; Section too long for Excel.

R	S	T	U	V	W	X	Y	Z
<b>7A. Market Risk</b>								
Corporate Governance and Stockholder Matters; Certain Relationships								

Cells pertaining to Named Entity Recognition will contain a bracketed, comma-separated list of named entities. If you did not select the Named Entity Recognition checkbox in Step 3, cells under the NER columns will be blank.

## Errors

### Error Dialog

Most common errors will be caught by the app, and displayed in a Red dialog box.

cess



## SEC EDGAR 10-K Information Access

FORD MOTOR CO (F, F-PC, F-PB)

Start Date: 5/4/2021

End Date: 5/4/2023

Form Type:

Read from file (.txt):

sample\_upload.txt

End date cannot be later than today's date

Entity Name	CIK Number	Form Type	Filing Date	Link to Document	Extract Info?
-------------	------------	-----------	-------------	------------------	---------------

For example, the end of the date range when searching for entities cannot be later than today's date.

Some other user-based errors which will be displayed here are the start date being later than the end date, or an uploaded CIK number being invalid.

cess



## SEC EDGAR 10-K Information Access

FORD MOTOR CO (F, F-PC, F-PB)

Start Date: 4/30/2022

End Date: 5/1/2022

Form Type:

Read from file (.txt):

sample\_upload.txt

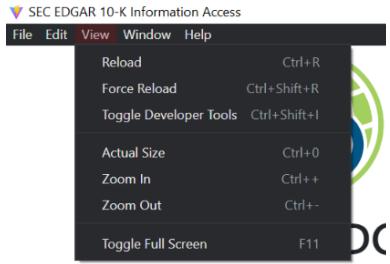
No filings found

Entity Name	CIK Number	Form Type	Filing Date	Link to Document	Extract Info?
-------------	------------	-----------	-------------	------------------	---------------

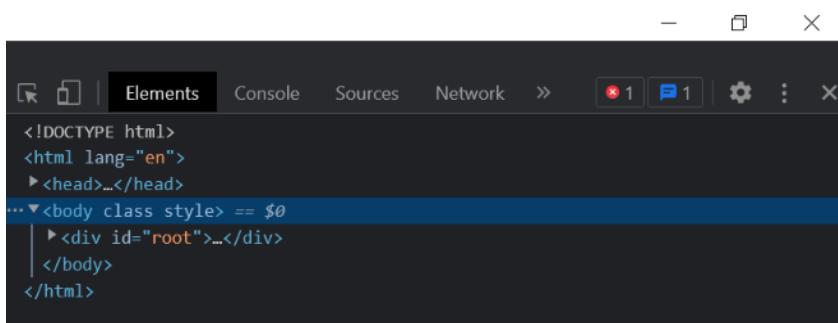
Note: if nothing is showing up in the results table, that means that there are no filings from that entity submitted to the SEC for that form type during the specified date range.

## Developer Tools

If you are experiencing errors and cannot tell why, you may be able to find out what is happening by viewing the developer tools.



On the top menu, click on the View tab.  
Click the Toggle Developer Tools option.



Click the Console tab.

```

Array(10)
searched
Array(10)
searched
Array(10)
searched
Array(10)
searched
Array(3)
CIK0000037996
2020-01-01
2022-03-28
Object
Array(3)
L
L
L
CIK0000047111
2019-12-31
2020-12-31
Object
Array(2)
L
L
C:\Users\chloe\laf_senior_project\extraction_results\react-dom.production.min.js:101
Uncaught TypeError: Cannot read properties of undefined (reading 'path')
    at An (C:\Users\chloe\laf_senior_project\extraction_results\react-dom.production.min.js:101:39)
    at Od (C:\Users\chloe\laf_senior_project\extraction_results\react-dom.production.min.js:52:317)
    at _e (C:\Users\chloe\laf_senior_project\extraction_results\react-dom.production.min.js:52:472)
    at __ (C:\Users\chloe\laf_senior_project\extraction_results\react-dom.production.min.js:153:15)
    at Ip (C:\Users\chloe\laf_senior_project\extraction_results\react-dom.production.min.js:100:68)
    at Tp (C:\Users\chloe\laf_senior_project\extraction_results\react-dom.production.min.js:101:380)
    at react-dom.production.min.js:111:190
    at cv (C:\Users\chloe\laf_senior_project\extraction_results\react-dom.production.min.js:292:189)
    at v_ (C:\Users\chloe\laf_senior_project\extraction_results\react-dom.production.min.js:195:57)
    at Iv (C:\Users\chloe\laf_senior_project\extraction_results\react-dom.production.min.js:105:469)
C:\Users\chloe\laf_senior_project\extraction_results\react-dom.production.min.js:359
NER: true
App.js:341
> |

```

This will display statements that get written out to the application's console, including uncaught error messages.