
Python SEC Edgar Documentation

Release 0.1.0

Ryan S. McCoy

Sep 22, 2020

Contents

1	Python SEC Edgar	1
1.1	Features	1
1.2	Quick Start Guide	1
1.3	Configure Settings (Optional)	2
2	Alright, what did I just do?	7
2.1	Why download the Complete Submission Filing?	8
2.2	Central Index Key (CIK)	8
2.3	Accession Number	8
2.4	Filings Statistics	8
2.5	Download Time Estimates	9
3	Todo	11
4	Installation	13
4.1	From sources	13
5	Usage	15
6	Contributing	17
6.1	Types of Contributions	17
6.2	Get Started!	18
6.3	Pull Request Guidelines	19
6.4	Tips	19
6.5	Deploying	19
7	Credits	21
7.1	Development Lead	21
7.2	Contributors	21
8	History	23
8.1	0.1.0 (2018-05-30)	23
8.2	0.1.1 (2020-09-08)	23
9	Indices and tables	25

CHAPTER 1

Python SEC Edgar

A Python application used to download and parse complete submission filings from the sec.gov/edgar website. The goal for this project is to make it easy to get filings from the SEC website onto your computer for the companies and forms you desire.

A few hurdles that I've tried to ease with this project:

- CIK to Ticker Equivalent - probably the biggest hurdle is just figuring out the CIK for the company you want. I've tried to bypass this via a reference file mapping CIK to tickers. I'm sure there is a better way, but for now it seems to work.
- Organizing the Data - I decided to keep it simple and organize the data similar to the SEC Edgar website (which is explained below)

1.1 Features

- Filter by Ticker
- Filter by Form Type
- Extract contents of Complete Submission Filing

1.2 Quick Start Guide

1.2.1 Setup Environment (Windows)

```
git clone https://github.com/ryansmccoy/py-sec-edgar.git
cd py-sec-edgar
conda create -n py-sec-edgar python=3.8 pandas numpy lxml -y
activate py-sec-edgar
pip install -r requirements.txt
```

1.2.2 Setup Environment (Linux):

```
git clone https://github.com/ryansmccoy/py-sec-edgar.git
cd py-sec-edgar
conda create -n py-sec-edgar python=3.8 pandas numpy lxml -y
source activate py-sec-edgar
sudo mkdir /sec_gov
sudo chown -R $USER:$USER /sec_gov
pip install -r requirements.txt
```

1.3 Configure Settings (Optional)

py-sec-edgar/py_sec_edgar/settings.py

1.3.1 Extracting Contents from Complete Submission Filing:

```
# extract all contents from txt file
# Set this to True and everything will be extracted from Complete Submission Filing
# Note: There is a lot of content in these filings, so be prepared

extract_filing_contents = False
```

1.3.2 Specify Form Types, Start, and End Dates:

```
# complete list @ py-sec-edgar/refdata/filing_types.xlsx

forms_list = ['10-K', '20-F']

# the urls of all filings are stored in index files
# so need to download these index files
# below just says download all of them

start_date = "1/1/2018"
end_date = "1/1/2025"
```

1.3.3 Specify Tickers:

```
# py-sec-edgar/refdata/tickers.csv

AAPL
MSFT
XOM
GOOGL
WFC
```

1.3.4 Run Application

```
$ cd py-sec-edgar
$ python py_sec_edgar
```

Above, is the same as running (See notes at top of `__main__.py` file for explanation):

```
$ cd py-sec-edgar
$ python py_sec_edgar/__main__.py
```

OR

```
$ cd py-sec-edgar
$ python py_sec_edgar/example.py
```

1.3.5 Output:

```
Starting Index Download:

Downloading Latest https://www.sec.gov/Archives/edgar/full-index/master.idx

Downloading:      https://www.sec.gov/Archives/edgar/full-index/master.idx
Saving to:  C:\sec_gov\Archives\edgar\full-index\master.idx
Selected User-Agent:      {'User-Agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_
↳12_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/54.0.2840.98 Safari/537.36'}
Success!      Saved to filepath:      C:\sec_gov\Archives\edgar\full-index\master.idx

      Completed Index Download
CIK                                72971
Company Name                      WELLS FARGO & COMPANY/MN
Form Type                        10-K
Date Filed                       2019-02-27
Filename                         edgar/data/72971/0000072971-19-000227.txt
published                        2019-02-27
url                              https://www.sec.gov/Archives/edgar/data/72971/...
Name: 103670, dtype: object
2019-05-01 14:14:49,841 ERROR py_sec_edgar.filing: Filing Already Exists
2019-05-01 14:14:51,844 INFO py_sec_edgar.filing: Filing Loaded
2019-05-01 14:14:55,613 INFO py_sec_edgar.filing: Filing Lxml

      GROUP                                KEY
↳VALUE
1      COMPANY DATA      0000072971-19-000227.hdr.sgml
↳20190227
2      COMPANY DATA      <acceptance-datetime>
↳20190227152351
4      COMPANY DATA      ACCESSION NUMBER      0000072971-19-
↳000227
5      COMPANY DATA      CONFORMED SUBMISSION TYPE
↳10-K
6      COMPANY DATA      PUBLIC DOCUMENT COUNT
↳ 211
7      COMPANY DATA      CONFORMED PERIOD OF REPORT
↳20181231
8      COMPANY DATA      FILED AS OF DATE
↳20190227
```

(continues on next page)

(continued from previous page)

```

9      COMPANY DATA          DATE AS OF CHANGE          _
↳20190227
14     COMPANY DATA          COMPANY CONFORMED NAME        WELLS FARGO & _
↳COMPANY/MN
15     COMPANY DATA          CENTRAL INDEX KEY              _
↳0000072971
16     COMPANY DATA  STANDARD INDUSTRIAL CLASSIFICATION  NATIONAL COMMERCIAL BANKS _
↳[6021]
17     COMPANY DATA          IRS NUMBER                      _
↳410449260
18     COMPANY DATA          STATE OF INCORPORATION          _
↳ DE
19     COMPANY DATA          FISCAL YEAR END                 _
↳1231
22     FILING VALUES          FORM TYPE                _
↳10-K
23     FILING VALUES          SEC ACT                      _
↳1934 Act
24     FILING VALUES          SEC FILE NUMBER        001-
↳02979
25     FILING VALUES          FILM NUMBER            _
↳19637386
28     BUSINESS ADDRESS        STREET 1                420 MONTGOMERY _
↳STREET
29     BUSINESS ADDRESS        CITY                    SAN _
↳FRANCISCO
30     BUSINESS ADDRESS        STATE                   _
↳ CA
31     BUSINESS ADDRESS        ZIP                      _
↳94163
32     BUSINESS ADDRESS        BUSINESS PHONE          _
↳6126671234
35     MAIL ADDRESS            STREET 1                420 MONTGOMERY _
↳STREET
36     MAIL ADDRESS            CITY                    SAN _
↳FRANCISCO
37     MAIL ADDRESS            STATE                   _
↳ CA
38     MAIL ADDRESS            ZIP                      _
↳94163
41     FORMER COMPANY          FORMER CONFORMED NAME    WELLS FARGO & _
↳CO/MN
42     FORMER COMPANY          DATE OF NAME CHANGE      _
↳19981103
45     FORMER COMPANY          FORMER CONFORMED NAME    NORWEST _
↳CORP
46     FORMER COMPANY          DATE OF NAME CHANGE      _
↳19920703
49     FORMER COMPANY          FORMER CONFORMED NAME    NORTHWEST _
↳BANCORPORATION
50     FORMER COMPANY          DATE OF NAME CHANGE      _
↳19830516
51     FORMER COMPANY          </acceptance-datetime>
2019-05-01 14:14:59,984 INFO py_sec_edgar.filing:

```

Extracting Filing Documents:

(continues on next page)

(continued from previous page)

```

2019-05-01 14:15:07,547 INFO py_sec_edgar.filing:
  TYPE SEQUENCE                                FILENAME
  ↪          RELATIVE_FILEPATH                DESCRIPTION
1      wfc-12312018x10k.htm          10-K          1
  ↪          FORM 10-K  000007297119000227\0001-(10...  0001-(10-K)_FORM_10-K_
  ↪wfc-12312018x10k.htm
2      wfc-12312018xex10a.htm        EX-10.A          2
  ↪          EXHIBIT 10.A  000007297119000227\0002-(EX...  0002-(EX-10.A)_EXHIBIT_10.A_
  ↪wfc-12312018xex10a...
3      wfc-12312018xex10c.htm        EX-10.C          3
  ↪          EXHIBIT 10.C  000007297119000227\0003-(EX...  0003-(EX-10.C)_EXHIBIT_10.C_
  ↪wfc-12312018xex10c...
4      wfc-12312018xex10i.htm        EX-10.I          4
  ↪          EXHIBIT 10.I  000007297119000227\0004-(EX...  0004-(EX-10.I)_EXHIBIT_10.I_
  ↪wfc-12312018xex10i...
5      wfc-12312018xex10j.htm        EX-10.J          5
  ↪          EXHIBIT 10.J  000007297119000227\0005-(EX...  0005-(EX-10.J)_EXHIBIT_10.J_
  ↪wfc-12312018xex10j...
204      R183.htm                    XML          204
  ↪ IDEA: XBRL DOCUMENT  000007297119000227\0204-(XM...  0204-(XML)_IDEA_
  ↪XBRL_DOCUMENT_R183.htm
205      R184.htm                    XML          205
  ↪ IDEA: XBRL DOCUMENT  000007297119000227\0205-(XM...  0205-(XML)_IDEA_
  ↪XBRL_DOCUMENT_R184.htm
206      R185.htm                    XML          206
  ↪ IDEA: XBRL DOCUMENT  000007297119000227\0206-(XM...  0206-(XML)_IDEA_
  ↪XBRL_DOCUMENT_R185.htm
207      Financial_Report.xlsx        EXCEL          207
  ↪ IDEA: XBRL DOCUMENT  000007297119000227\00000729...
  ↪Financial_Report.xlsx
208      Show.js                    XML          208
  ↪ IDEA: XBRL DOCUMENT  000007297119000227\0208-(XM...  0208-(XML)_IDEA_
  ↪XBRL_DOCUMENT_Show.js
209      report.css                  XML          209
  ↪ IDEA: XBRL DOCUMENT  000007297119000227\0209-(XM...  0209-(XML)_IDEA_XBRL_
  ↪DOCUMENT_report.css
210      FilingSummary.xml           XML          211
  ↪ IDEA: XBRL DOCUMENT  000007297119000227\0211-(XM...  0211-(XML)_IDEA_XBRL_
  ↪DOCUMENT_FilingSummary.xml
211  0000072971-19-000227-xbrl.zip    ZIP          213
  ↪ IDEA: XBRL DOCUMENT  000007297119000227\00000729...
  ↪0000072971-19-000227-xbrl.zip

[211 rows x 6 columns]
2019-05-01 14:15:07,690 INFO py_sec_edgar.filing:

Extraction Complete

```


CHAPTER 2

Alright, what did I just do?

- Created folder structure which mimics sec.gov website structure (see “Paths and Directory Structure” below)
- Downloaded the necessary idx files (files containing the links to the sec filings) and merge them into combined file (fyi 1.9 GB csv)
- loaded tickers from tickers.csv file and filters them
- load forms from settings.py file and filters them
- filter against forms set in settings.py file, and will then start downloading the individual filings for the tickers in the tickers.csv file.

Paths and Directory Structure

sec.gov website:

```
https://www.sec.gov/  
  
https://www.sec.gov/Archives/edgar/full-index/ <- path where "index" files are located  
  
https://www.sec.gov/Archives/edgar/full-index/2018/QTR1/master.idx <- EDGAR Index_  
↳ Files are tab delimited txt files  
  
https://www.sec.gov/Archives/edgar/data/ <- path where all the actual filings are_  
↳ stored  
  
https://www.sec.gov/Archives/edgar/data/1041588/0001041588-18-000005.txt <- these are_  
↳ the complete submission file  
  
https://www.sec.gov/Archives/edgar/data/<CIK>/<ACCESSION_NUMBER_WITHOUT_DASHES>/  
↳ <ACCESSION_NUMBER>.txt <- follows this format
```

local folder equivalent:

```
C:\sec_gov\
```

(continues on next page)

(continued from previous page)

```

C:\sec_gov\Archives\edgar\full-index\ <- path where "index" files are located

c:\sec_gov\Archives\edgar\full-index\2018\QTR1\master.idx <- EDGAR Index Files are_
↳tab delimited txt files

c:\sec_gov\Archives\edgar\data\ <- path where all the actual filings are stored

c:\sec_gov\Archives\edgar\data\1041588\000104158818000005\0001041588-18-000005.txt <-_
↳these are the complete submission file

c:\sec_gov\Archives\edgar\data\<CIK>\<ACCESSION_NUMBER_WITHOUT_DASHES>\<ACCESSION_
↳NUMBER>.txt <- follow this format

```

2.1 Why download the Complete Submission Filing?

- **Most Efficient and Courteous way of getting data from SEC website**
 - Contains everything the company filed in filing in one file
 - Not making multiple download requests per filing

2.2 Central Index Key (CIK)

The CIK is the unique numerical identifier assigned by the EDGAR system to filers when they sign up to make filings to the SEC. CIK numbers remain unique to the filer; they are not recycled.

2.3 Accession Number

In the example above, “0001193125-15-118890” is the “accession number,” a unique identifier assigned automatically to an accepted submission by the EDGAR Filer System. The first set of numbers (0001193125) is the CIK of the entity submitting the filing. This could be the company or a third-party filer agent. Some filer agents without a regulatory requirement to make disclosure filings with the SEC have a CIK but no searchable presence in the public EDGAR database. The next 2 numbers (15) represent the year. The last series of numbers represent a sequential count of submitted filings from that CIK. The count is usually, but not always, reset to 0 at the start of each calendar year.

2.4 Filings Statistics

Form 4	6,420,154	(Ownership)
8-K	1,473,193	(Press Releases)
10-K	180,787	(Annual Report)
10-Q	552,059	(Quarterly Report)
13F-HR	224,996	(Investment Fund Holdings)
S-1	21,366	(IPO offering)

Total	17,492,303	

2.5 Download Time Estimates

```
180,787      10-K filings
      8      seconds on average to download single filing
-----
1,446,296    seconds
24,104.93    minutes
401.75       hours
-----
16.74        days to download all 10-K filings via 1 connection
```


CHAPTER 3

Todo

- Feeds
 - Make Full-Index more efficient
 - Incorporate RSS Feed
- Add Multi-Threading
- need to figure out way to quickly access downloaded content
- extract earnings data from 8-K
- setup proper logging instead of print
- add tests
- need to add add way to quickly update new tickers

4.1 From sources

The sources for Python SEC Edgar Data can be downloaded from the [Github repo](#).

You can either clone the public repository:

```
$ git clone git://github.com/ryansmccoy/py_sec_edgar
```

Once you have a copy of the source, you can install it with:

```
$ pip install -r requirements.txt
```


CHAPTER 5

Usage

To use Python SEC Edgar Data in a project:

```
import py_sec_edgar
```


Contributions are welcome, and they are greatly appreciated! Every little bit helps, and credit will always be given. You can contribute in many ways:

6.1 Types of Contributions

6.1.1 Report Bugs

Report bugs at <https://github.com/ryansmccoy/py-sec-edgar/issues>.

If you are reporting a bug, please include:

- Your operating system name and version.
- Any details about your local setup that might be helpful in troubleshooting.
- Detailed steps to reproduce the bug.

6.1.2 Fix Bugs

Look through the GitHub issues for bugs. Anything tagged with “bug” and “help wanted” is open to whoever wants to implement it.

6.1.3 Implement Features

Look through the GitHub issues for features. Anything tagged with “enhancement” and “help wanted” is open to whoever wants to implement it.

6.1.4 Write Documentation

Python SEC Edgar Data could always use more documentation, whether as part of the official Python SEC Edgar Data docs, in docstrings, or even on the web in blog posts, articles, and such.

6.1.5 Submit Feedback

The best way to send feedback is to file an issue at <https://github.com/ryansmccoy/py-sec-edgar/issues>.

If you are proposing a feature:

- Explain in detail how it would work.
- Keep the scope as narrow as possible, to make it easier to implement.
- Remember that this is a volunteer-driven project, and that contributions are welcome :)

6.2 Get Started!

Ready to contribute? Here's how to set up *py-sec-edgar* for local development.

1. Fork the *py-sec-edgar* repo on GitHub.
2. Clone your fork locally:

```
$ git clone git@github.com:your_name_here/py-sec-edgar.git
```

3. Install your local copy into a virtualenv. Assuming you have virtualenvwrapper installed, this is how you set up your fork for local development:

```
$ mkvirtualenv py-sec-edgar
$ cd py-sec-edgar/
$ python setup.py develop
```

4. Create a branch for local development:

```
$ git checkout -b name-of-your-bugfix-or-feature
```

Now you can make your changes locally.

5. When you're done making changes, check that your changes pass flake8 and the tests, including testing other Python versions with tox:

```
$ flake8 py-sec-edgar tests
$ python setup.py test or py.test
$ tox
```

To get flake8 and tox, just pip install them into your virtualenv.

6. Commit your changes and push your branch to GitHub:

```
$ git add .
$ git commit -m "Your detailed description of your changes."
$ git push origin name-of-your-bugfix-or-feature
```

7. Submit a pull request through the GitHub website.

6.3 Pull Request Guidelines

Before you submit a pull request, check that it meets these guidelines:

1. The pull request should include tests.
2. If the pull request adds functionality, the docs should be updated. Put your new functionality into a function with a docstring, and add the feature to the list in README.rst.
3. The pull request should work for Python 2.7, 3.4, 3.5 and 3.6, and for PyPy. Check https://travis-ci.org/ryansmccoy/py-sec-edgar/pull_requests and make sure that the tests pass for all supported Python versions.

6.4 Tips

To run a subset of tests:

```
$ py.test tests.test_py-sec-edgar
```

6.5 Deploying

A reminder for the maintainers on how to deploy. Make sure all your changes are committed (including an entry in HISTORY.rst). Then run:

```
$ bumpversion patch # possible: major / minor / patch
$ git push
$ git push --tags
```

Travis will then deploy to PyPI if tests pass.

7.1 Development Lead

- Ryan S. McCoy <github@ryansmccoy.com>

7.2 Contributors

None yet. Why not be the first?

8.1 0.1.0 (2018-05-30)

- First release.

8.2 0.1.1 (2020-09-08)

- Refactored.

CHAPTER 9

Indices and tables

- `genindex`
- `modindex`
- `search`