# SEC 10-Q Filings extraction & Scraping web data from URL.

## Project Overview

The **SEC Filing app** is a Streamlit-based web application that allows users to interact with SEC filings data. It provides functionalities for extracting, filtering, and analyzing 10-Q filings from the U.S. Securities and Exchange Commission's website. Additionally, users can extract text from SEC filings and analyze sections with the help of AI. The webapp also allows users to access the underlying code files and documentation.

---

# Key Features

## Task 1: 10-Q Filings

- Fetch and display SEC 10-Q filings for specific years and quarters.
- Filter filings based on various attributes like CIK, Form Type, and Date.
- Option to download the results as a CSV or text file.

## Task 2: URL Text Extraction

- Allows users to extract specific sections from an SEC filing URL.
- Option to specify the start and end sections to extract.
- After extraction, AI processing is applied to analyze the extracted content.

## Task 3: Combined task1 and task2 with UI hosted on cloud streamlit.

## Code Files & Documentation

- Displays code files related to the SEC Filing Analyzer.
- Option to view Python code files with syntax highlighting.
- Option to download documentation in PDF format.
- Code files included:
    - combined_tsk1_tsk2_with_ui.py (307 lines)
    - task1.py (95 lines)
    - task2.py (66 lines)
    - Documentation.pdf

**Time Taken for Task1.py Task2.py -** 45 minutes

**Time Taken for Combined with UI:** 3 hrs

# Task 1: Fetching 10-Q Filings

## File Name: task1.py

## Description:

This script downloads and extracts 10-Q filings for a given year and quarter from the SEC EDGAR database.

## Dependencies:

- requests
- time
- re
- csv
- sys

## Usage:

python task1.py <YEAR> <QUARTER>

Example:

python task1.py 2024 2

## Code Summary:

- **download_crawler_index(year, quarter)**: Downloads the crawler.idx file from the SEC website.
- **extract_10Q_filings(data, year, quarter)**: Parses and extracts 10-Q filings from the downloaded data.
- **Command-line execution (__main__)**: Reads input arguments (year & quarter), validates them, and executes the functions.

## Output:

- A CSV file named 10Q_filings_<YEAR>_Q<QUARTER>.csv containing extracted 10-Q filing details.

## Task 2: URL Text Extraction

**File Name:** task2.py

**Description:**

This script extracts specific sections from an SEC filing document based on a given URL. Users can optionally specify start and end section names to refine the extraction.

**Dependencies:**

- requests
- re
- sys
- BeautifulSoup (from bs4)

**Code Summary:**

- fetch_document(url): Downloads the SEC filing document from the given URL.
- extract_section_text(html, start_section, end_section): Parses the HTML using BeautifulSoup and extracts text between the specified start and end sections.
- **Command-line execution (__main__)**:
    - Reads the SEC filing URL from the command line.
    - Calls fetch_document() to retrieve the HTML content.
    - Calls extract_section_text() to extract and display relevant content.

**Output:**

- Extracted text from the SEC filing is printed to the console.
- If sections are specified, only the relevant part of the document is shown.

## Combined task1 and task2 with ui

Technical Components

1. SEC Data Retrieval Uses the SEC's public archives (crawler.idx files) to find 10-Q filings Implements proper headers to avoid 403 Forbidden errors Parses filing information into a structured DataFrame

2. Document Processing Handles both HTML and plain text filings Extracts specific sections using regex pattern matching Cleans HTML content by removing scripts, styles, and navigation elements

3. AI Integration Connects to Groq's API using the llama-3.3-70b-versatile model Processes extracted text for analysis Handles API errors and timeouts gracefully

4. User Interface Streamlit-based responsive interface Organized into three main tasks with clear navigation Includes download options for all results.

**FAQ'S**

1. What does the SEC Filing Analyzer do?

The SEC Filing Analyzer helps users fetch, filter, and download SEC 10-Q filings for specific years and quarters. It also allows users to extract sections from SEC filings using a URL and perform AI-based analysis on the extracted content. Additionally, it provides access to the project's source code and documentation.

2. How do I use the SEC Filing Analyzer? Step 1: Choose a task from the sidebar: Task 1: Fetch 10-Q Filings: Specify a year and select the quarters you want to analyze. Task 2: URL Text Extraction: Provide the URL of an SEC filing and specify the section to extract. Task 3: Code Files & Documentation: View the code files or download the project documentation. Step 2: Follow the prompts for each task and view the results. Step 3: Use the available download options to get results in CSV or text formats.

3. What are the inputs required for Task 1: 10-Q Filings?

Year: Enter a year (1995-2025). Quarters: Select the quarters for which you want to fetch the 10-Q filings (1, 2, 3, or 4). Click "Fetch Filings" to view the results.

4. What are the inputs required for Task 2: URL Text Extraction? SEC Filing URL: Paste the URL of the SEC filing you want to extract. Start Section: Optionally specify the start section of the document to extract Ex:"Filing Date". End Section: Optionally specify the end section of the document to extract Ex:"Data Files". Click "Extract Section" to begin the extraction.

5. How can I access the code and documentation for the project (Task 3)?

In Task 3, you can select the desired code file or the documentation.pdf. For Python files, the code will be displayed directly within the app. For the PDF documentation, a download button will be provided.