

Requerimiento #5 - Publicaciones de Estados Financieros

En su quinta semana de trabajo, su jefe le pide colaboración para perfeccionar los procesos de publicación de los Estados Financieros de su compañía. En particular, su jefe le comenta que la empresa nunca se ha preocupado de llevar registro del tiempo que les toma elaborar los Estados Financieros trimestre a trimestre, y que desean comenzar a hacerlo. Debido a esto, usted tendrá que partir recolectando la poca información útil que se encuentra disponible en internet.

El Requerimiento

Buenos días,

Producto de un acuerdo entre las áreas de Contabilidad, Recursos Humanos y Operaciones, se ha decidido crear una base de datos que lleve registro del tiempo que toma construir los Estados Financieros. La idea es que más adelante podamos mejorar nuestros procesos y que ahorremos recursos que se podrían destinar a otros fines. Para poder comenzar a detectar ineficiencias, te pedimos que comiences a armar la base de datos.

La tabla de la base de datos debería verse así:

<i>Trimestre_Estado_Financiero</i>	<i>Fecha_Primer_envio</i>	<i>Fecha_ultimo_envio</i>
<i>31-03-2020</i>	<i>15-04-2020</i>	<i>03-05-2020</i>
<i>30-06-2020</i>	<i>01-08-2020</i>	<i>01-08-2020</i>
<i>30-09-2020</i>	<i>03-12-2021</i>	<i>11-02-2021</i>
<i>...</i>	<i>...</i>	<i>...</i>

En el link de más abajo podrás acceder a todos los Estados Financieros publicados para el primer semestre de 2020, para cada una de las compañías fiscalizadas por la CMF. Entre ellas, la nuestra. El link es el siguiente:

https://www.cmfchile.cl/institucional/mercados/novedades_envio_sa_ifrs.php?mm_ifrs=03&aa_ifrs=2020

Por dónde partir...

Al terminar de leer el correo, usted se dirige al escritorio de su colega que tiene más experiencia. Su colega, lamentablemente, le señala que está en otro proyecto y que no puede ayudarlo mucho. Sin embargo, le dice que lo lógico que habría que hacer sería armar un código que haga webscraping de la web de la CMF, cambiando las fechas de la URL, y guardando esa información en la base SQL. Le recuerda que, para esto último, Pandas tiene una función bien útil llamada “to_sql”.

Respecto a la web de la CMF, luego de una rápida mirada, su colega le recomienda que extraiga la información directamente desde el HTML que entrega la web. Para extraer la información que se requiere, le recomienda utilizar la librería “Beautiful Soup”, la cual permite, por ejemplo, extraer información de ciertas etiquetas específicas, como la etiqueta “<tr>”.

Le indica además que lo más importante es poder tener pronto la base de datos para probarla en los otros procesos que se generarán a partir de ella, y que los códigos desarrollados (al menos los más relevantes) puedan estar en un archivo .py.

- **La tabla de datos con la información debe llamarse “FECHAS_EEFF”.**
- **Llame al archivo de la base de datos “datos_eeff.db”.**
- **Llame al archivo con códigos “auxiliares5.py”.**

Finalmente, le da algunos consejos sobre webscraping:

- Evitar lo más posible hacer consultas repetitivas a una misma página web.
 - Al estar desarrollando el código para extraer la información de la web, asegurarse que no se consulte una y otra vez la misma página.
- Evitar hacer consultas muy juntas una de otra.
 - Para ello se puede usar una librería de Python llamada “time”, que permite pausar la ejecución de un código (time.sleep(5)).

La Entrega

Fecha de entrega: 30 de noviembre (hasta las 23.59hrs).

Entregables:

- Archivo auxiliares5.py
- Archivo datos_eeff.db

Todos los entregables deben estar en una sub-carpeta de la carpeta compartida llamada “requerimiento5”.