import requests

import html

import pandas as pd

from bs4 import BeautifulSoup

def get\_stop\_info(stop\_link: str) -> dict:

    url = f'https://pda5284.gov.taipei/MQS/{stop\_link}'

    # get url and save to html file

    # the html file is saved files as bus\_{stop\_link}.html

    response = requests.get(url)

    if response.status\_code == 200:

        # read id from url

        stop\_id = stop\_link.split("=")[1]

        with open(f"bus\_stop\_{stop\_id}.html", "w", encoding="utf-8") as file:

            file.write(response.text)

        print(f"網頁已成功下載並儲存為 bus\_{stop\_link}.html")

    else:

        print(f"無法下載網頁，HTTP 狀態碼: {response.status\_code}")

def get\_bus\_route(rid):

    """

    Retrieve two DataFrames containing bus stop names and their corresponding URLs based on the route ID (rid).

    Args:

        rid (str): Bus route ID.

    Returns:

        tuple: Two Pandas DataFrames, each corresponding to one direction of the bus route.

    Raises:

        ValueError: If the webpage cannot be downloaded or if insufficient table data is found.

    """

    url = f'https://pda5284.gov.taipei/MQS/route.jsp?rid={rid}'

    # Send GET request

    response = requests.get(url)

    # Ensure the request is successful

    if response.status\_code == 200:

        # Parse HTML using BeautifulSoup

        soup = BeautifulSoup(response.text, "html.parser")

        # Find all tables

        tables = soup.find\_all("table")

        # Initialize DataFrame list

        dataframes = []

        # Iterate through tables

        for table in tables:

            rows = []

            # Find all tr tags with the specified classes

            for tr in table.find\_all("tr", class\_=["ttego1", "ttego2", "ttback1", "ttback2"]):

                # Extract stop name and link

                td = tr.find("td")

                if td:

                    stop\_name = html.unescape(td.text.strip())  # Decode stop name

                    stop\_link = td.find("a")["href"] if td.find("a") else None

                    if stop\_link:

                        # Call get\_stop\_info function to get stop information

                        get\_stop\_info(stop\_link)

                    # Append to rows

                    rows.append({"stop\_name": stop\_name, "stop\_link": stop\_link})

            # If data exists, convert to DataFrame

            if rows:

                df = pd.DataFrame(rows)

                dataframes.append(df)

        # Return two DataFrames

        if len(dataframes) >= 2:

            return dataframes[0], dataframes[1]

        else:

            raise ValueError("Insufficient table data found.")

    else:

        raise ValueError(f"Failed to download webpage. HTTP status code: {response.status\_code}")

# Test function

if \_\_name\_\_ == "\_\_main\_\_":

    rid = "10417"  # Test route ID

    try:

        df1, df2 = get\_bus\_route(rid)

        print("去程:")

        print(df1)

        print("\n回程:")

        print(df2)

    except ValueError as e:

        print(f"Error: {e}")