ガソリン料金のランキング検索システム

**1. プロジェクト概要**

　プロジェクト名: ガソリン価格収集スクリプト

　目的 :　ガソリンの価格を収集して、価格変動をトラッキングする

　対象サイト : <https://gogo.gs/ranking/>

2. 対象データ

　rank(順位)

　gas(ガソリンの種類)

　(現金or会員)

　price(ガソリンの価格)

　shop(店舗名)

　address(店舗の住所)

データ形式

        for name in names:

            rank = int(name.select\_one('div.rank').text)

            gas = gasoline\_type

            mtype = name.select\_one('td.price-td > div:last-of-type').text

            price = int(name.select\_one('div.price').text)

            shop = name.select\_one('a.shop-name').text

            address = name.select\_one('p.address').text

            d\_list.append({

                'rank' : rank,

                'type' : gas,

                'mtype' : mtype,

                'price' : price,

                'shop' : shop,

                'address' : address

            })

3. スクレイピングの技術的要件

　仕様するプログラミング言語: Python

　ライブラリやフレームワーク : 仕様するライブラリ(BeautifulSoup,Requests,Selenium,pandas,sqlalchemy,tkinter)

　データ取得方法

　　・HTMLの解析方法(動的サイトはSelenium、静的サイトをBeautifulsoupで行う)

実行環境

　Python 3.8.20

BeautifulSoup 4.12.3

Requests 2.32.3

　Selenium 3.141.0

pandas 2.0.3

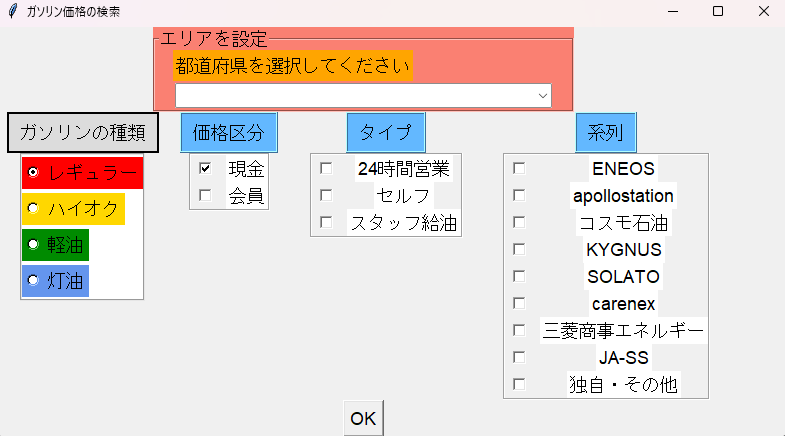
sqlalchemy 2.0.34

4. スクレイピングのフロー

　ステップ1: 対象となるエリアやタイプをTkinter(GUI)で選択

　　　　 　　 ・対象となるエリア、ガソリンの種類、価格区分、タイプ、系列を選択する

エリアを選択



タイプを選択

価格区分を選択

ガソリンの種類を選択

系列を選択

最後にOKボタンを押す

ステップ2: Tkinter(GUI)から選択された情報を取得

・ステップ1で選択した内容の情報を取得する

　ステップ3: 対象URLのリストを取得

・選択されたエリアの最大ページを取得する

　ステップ4: データの取得

・順位、ガソリンの価格、価格区分、店舗名、店舗住所を取得する

ガソリンの種類



ガソリンの価格

価格区分

店舗名

順位

最大ページ数

店舗住所

　ステップ5: データの保存

・収集したデータをPostgleSQLと連携し、データベースを保存する

　　　　　　　　　　　(CSVで保存する場合はCSVでコメントアウトしている部分を外す)



5. 注意点や制約事項

　対象サイトの利用規約: 対象サイトの利用規約にスクレイピングが禁止されていないことを

　　　　　　　　　　　　　　　　確認

　リクエストの頻度: サイトに過度な負荷をかけないよう、画面遷移時にsleep(3)を使い、

　　　　　　　　　　　　　待機時間を適切に設定

6. テストと検証

　プログラミングが正しく機能しているか、また収集したデータが正しいかを確認するため、

　別途テスト仕様書(ガソリン価格検索(テスト仕様書).xlsx)を作成し、検証

7. ポートフォリオ作成に伴い取り組んだこと

　　・言語がPythonなので、まずはPythonの基本的な操作(javaと比較しながら)を学習

　　・スクレピングの基礎を勉強するためにUdemyにて講義を受講

　　・GUIを作成したいと思い、Tkinterの作成方法を学習

　　・PythonとPostgreSQLを連携させるためのライブラリ(sqlalchemy)の扱い方を

　　　学習

8. 今後の改善点

　ファイルを一つにまとめてしまっているので、クラス分けなどをし、

　コードを見やすくする