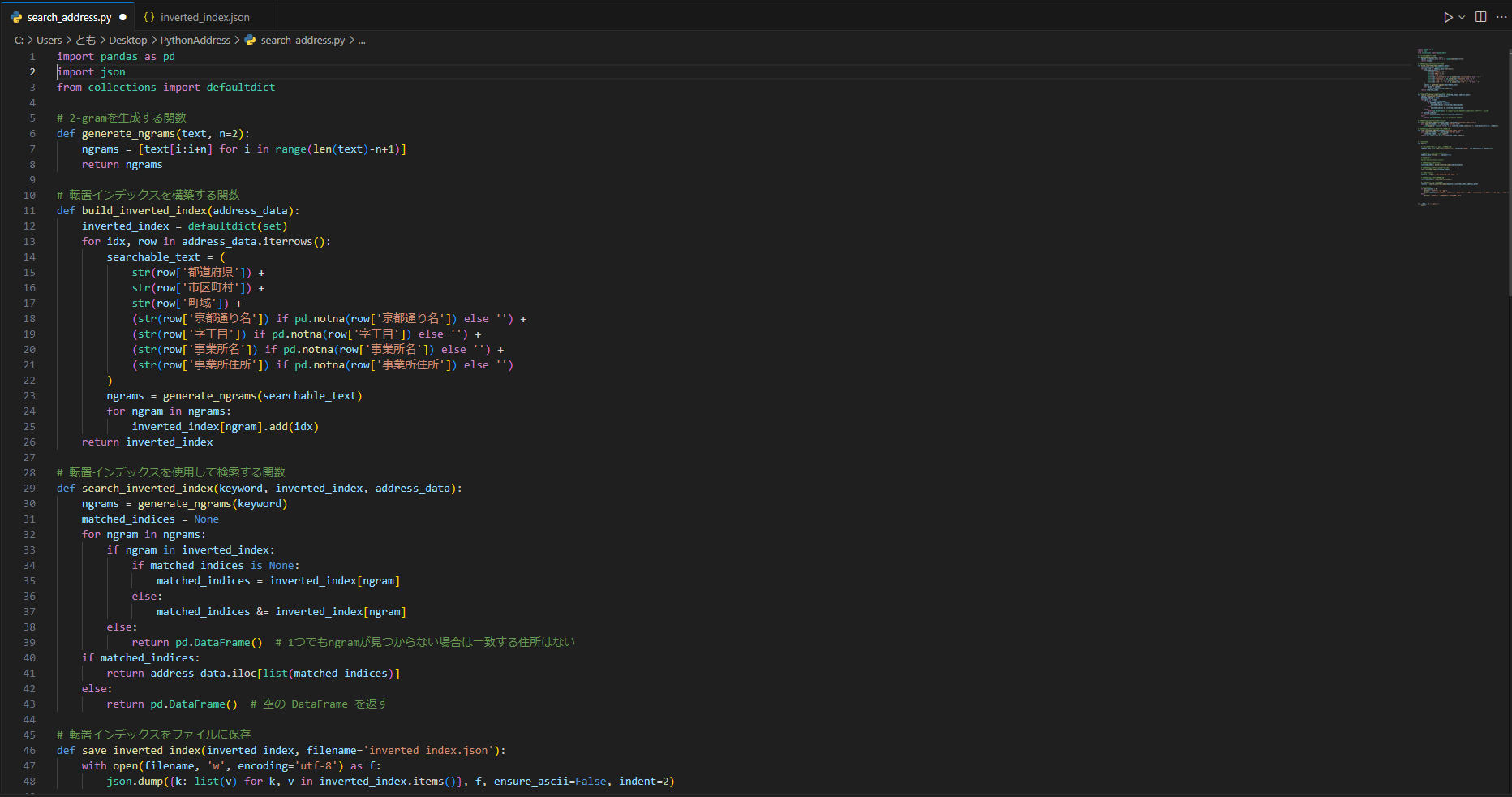
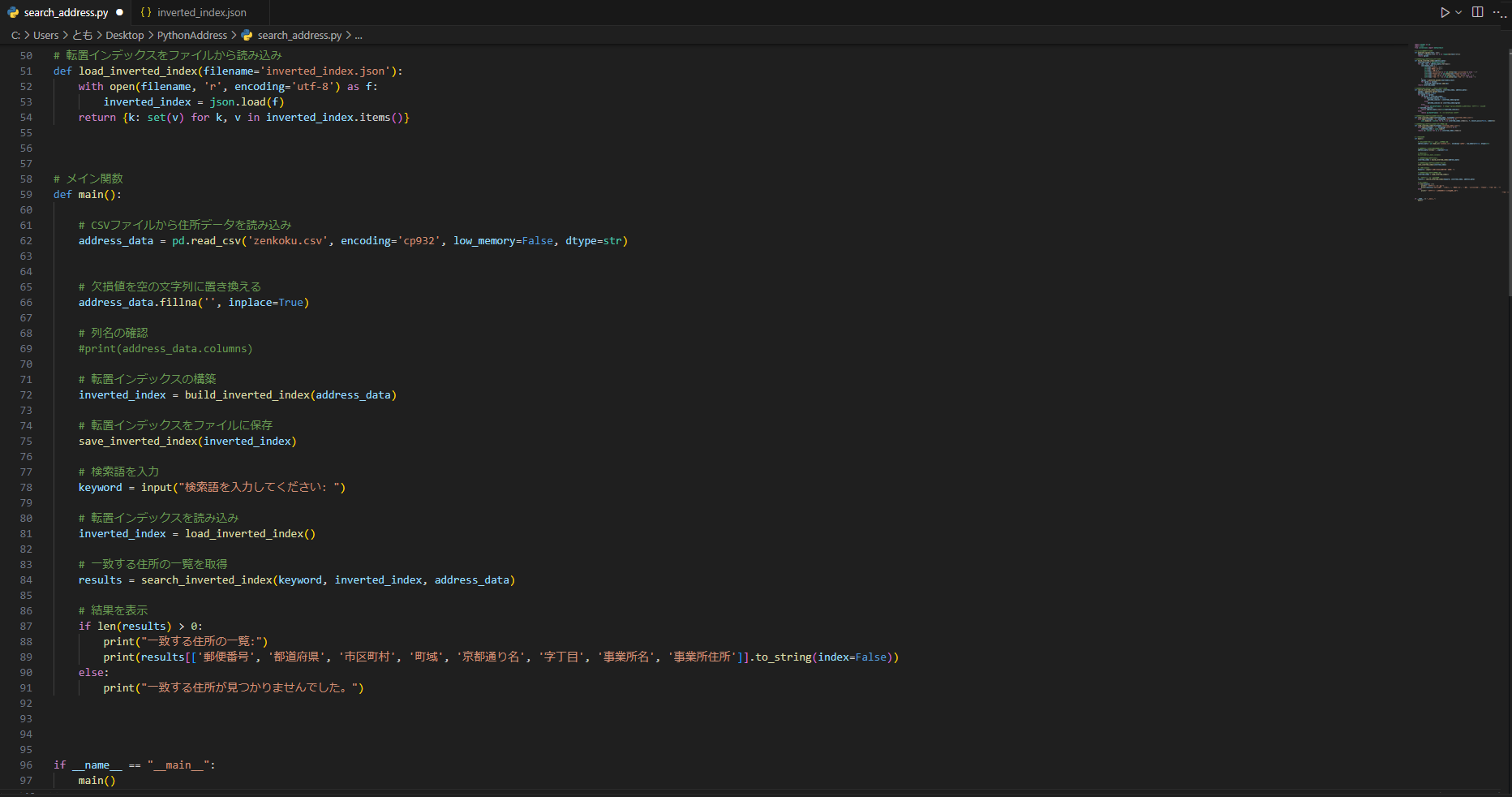
住所検索機能の技術説明資料

1. システムの全体図

言語：Python 　　エディタ：VScode

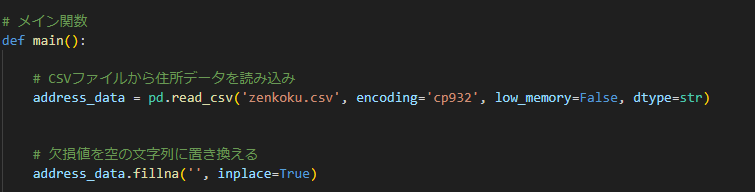
Pythonライブラリのpandasで全国の住所のcsvを取り込み、ユーザーが入力した検索ワードに対し2-gramにて転置インデックスを構築しそれをもとに部分一致検索を高速に効率よく出力するシステムです。各機能を関数で独立させているため保守性も高くなっています。





1. システム説明

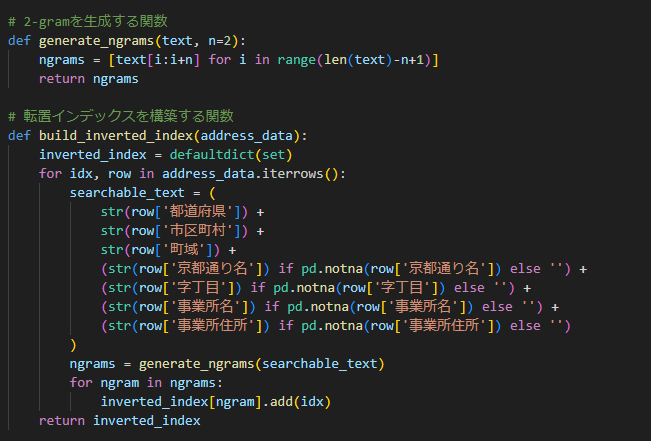
* Pandasによるcsvファイルの取り込み



Pandasライブラリのpd.read\_csv()メソッドにてcsvファイルを取得しています。なにもない値があればそれを空文字列に変換することによってエラーを回避しています。

* 転置インデックスの構築

inverted\_index = build\_inverted\_index(address\_data)



取り込んだcsvファイルをもとに2-gramの文字列を生成しそれらが出現するインデックスをすべて記述する転置インデックスを構築します。その際defaultdictを転置インデックスに代入することで都度デフォルト値を指定する必要がなくなります。

また、notnaメソッドでNaN値をチェックし、NaNなら空の文字列を返すようにしています。

save\_inverted\_index(inverted\_index)

モニター画面に映る文字

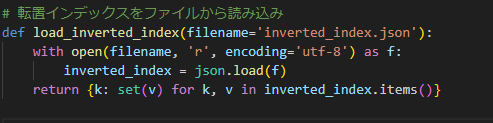
自動的に生成された説明

構築した転置インデックスをJSON形式でファイルに保存します。

これにより後々プログラムを再実行する際にインデックスを再構築しなおす必要をなくしています。

* 一致する住所の一覧を取得

inverted\_index = load\_inverted\_index()

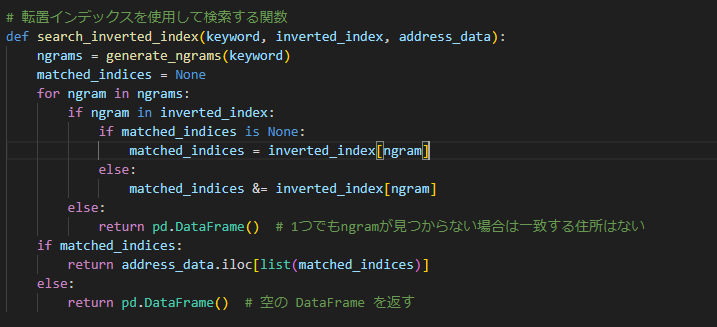


先ほどの転置インデックスが記載されたjsonファイルを読み込みます。

そして、読み込んだ転置インデックスをPythonの辞書型として返します。

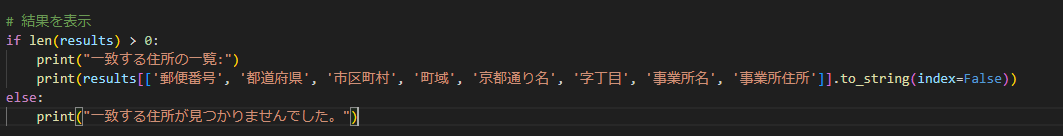
results = search\_inverted\_index(keyword, inverted\_index, address\_data)

読み込んだ転置インデックスファイル、ユーザーが入力した検索文字、全国住所csvを引数に検索文字と一致する住所の一覧を取得します。



検索キーワードから2-gramを生成し転置インデックスを部分一致検索します。その際2-gramを利用することで、入力されたキーワードが住所データ内のどの部分に現れても一致を見つけることができます。一致したインデックスが存在する場合、そのインデックスに対応する住所データを返します。存在しなければ空の値を返します。

* 結果の表示



検索結果を表示します。結果がある場合は住所の詳細を表示し結果がない場合は「一致する住所が見つかりませんでした」と表示します。