Spectrum Data Collection

ECE1551: Mobile Broadband Radio Access Networks

Navaneetha Krishna Madan Gopal

1005171127

Abstract

With the initial implementations of 5G in Canada around the corner, it is of great importance for providers to analyze various aspects of the spectrum. Data regarding the bands owned by operators and possible carrier aggregation scenarios will help tailor the solutions deployed to achieve better data rates. Data regarding the population and area covered, tiers spanned will help providers optimize their deployment solutions to obtain optimum utility of the resources deployed from an economic standpoint. Possible partnerships with other providers can also be identified through the analysis of this spectrum data.

Collecting all this data manually is a daunting and wasteful task. This project is a solution to automate the process of collecting data regarding the licenses held by various operators in different categories. Besides this, the frequency ranges, bandwidth and the population covered by the operator in each of the licences can also be obtained through this code.

Programming environment

The process of extracting data from websites called *web scraping* has been implemented using Python 3.7.1 in the Jupyter Notebook environment. The libraries being used are:

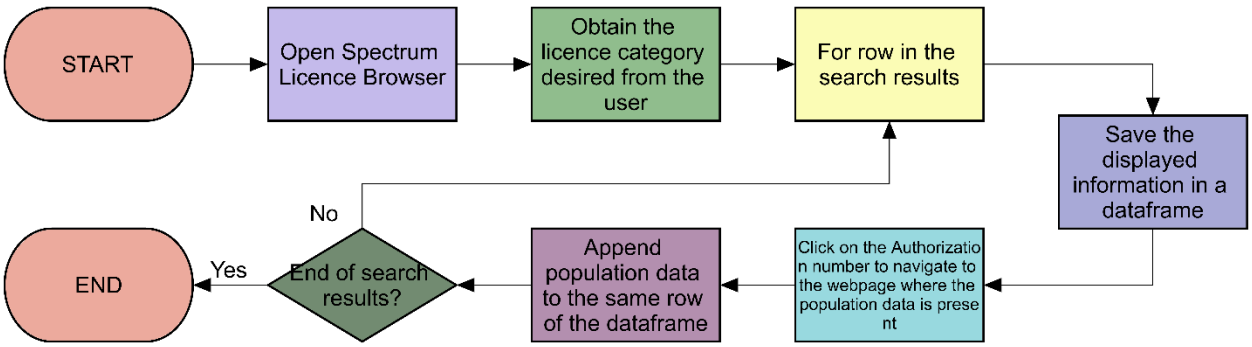
- selenium - used to control the opening, clicking and navigation in the web browser. In this case, Google Chrome is being used
- BeautifulSoup - used to extract information from the contents displayed on the webpage
- pandas - used to handle the extracted data and convert into convenient csv format
- tabula - used to obtain data from tables in the downloaded PDF files
- os - used to store, delete and perform other directory related tasks on the system running the code

Program flow

The program flow can be divided into two pipelines. The first pipeline extracts all the licences in the user specified licence category from the Spectrum Licence Browser. The data collected in this pipeline is:

- Authorization Number
- Former Authorization Number
- Company Name
- Account Number
- Licence Category
- Area Code (Tier)
- Area Name
- Subservice
- Population

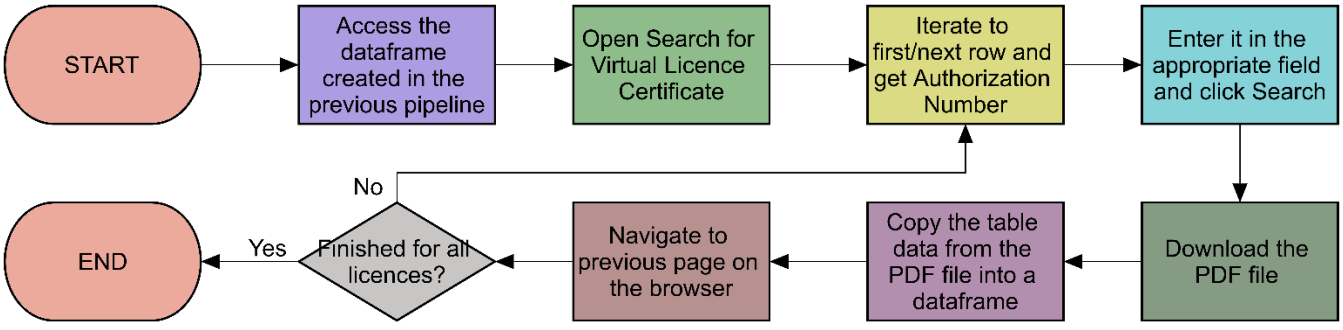
This is shown in the following flowchart:



The second pipeline extracts data regarding each of the licences. PDF files for each licence are downloaded from [Search for Virtual Licence or Certificate]. From the downloaded PDF file, the following data is extracted:

- Effective date
- Expiry date
- Licence number
- Account number
- Service type
- Licence holder type
- Frequency Ranges
- Spectrum
- Conditions

This is shown in the following flowchart:



Special features

One common problem with web scraping is the speed of the internet connection used to perform the tasks. If the connection is slow, the code will raise errors about not finding certain elements in the webpage as it has still noit been loaded. This can cause the entire data extraction pipeline to creash after a considerable amount of work has been done. This problem has been solved by incorporaing a "wait till button is loaded" feature thus increasing the reliability of this solution.

The option to also keep or delete the downloaded licence PDF files is provided to avoid unwanted clutter and memory consumption on the device.

Instructions for use

This code can be run on performing the following steps:

- Install the libraries listed above
- Download the driver of the browser being used and make the required changes in line 19
- Update the paths where you'd like to save the csv files and downloaded reports in lines 16 and 15

You are good to go!