GGE6505 Introduction to Big Data & Data Science

Assignment - 3 : Getting Data with APIs

Group 9

> Sonal Joshi (3719369)
>
> Iaswer Karthikeyan (3736196)
>
> Kishankumar Sachani (3740388)

```
# Installing packages which are not present by default
!pip install praw
```

(Que 1) We chose OpenStreetMap API.

- OpenStreetMap (OSM) is a free and open-source map of the world which contains geographic data such as roads, buildings, etc.
- The OpenStreetMap API is a web service that provides read and write access to the OSM data.
- Overpass API is a read-only API that allows users to query OSM data in a more flexible and efficient way than the OSM API.
- Overpass API provides a way to extract a custom selection of data from OSM based on different criteria like geographical region, tags, date/time, and much more. Overpass API supports multiple output formats, including XML, JSON, and GeoJSON.

```
# Importing the required packages
import requests
import json
import geojson
```

(Que 2) We will get the details about road network, buildings and other features for New Brunswick using Overpass API and query parameters.

> Access is public for reading or querying data using overpass, so there is no need for any API key. To modify data, an account would be required to do so.

```
# Define the API URL
url = "http://overpass-api.de/api/interpreter"
```

```
# Define the query to obtain details about the road network, buildings, and other features in New Brunswick
query = """
[out:json];
area["name"="New Brunswick"]->.boundaryarea;
(
  way["highway"](area.boundaryarea);
  relation["highway"](area.boundaryarea);
  way["building"](area.boundaryarea);
  relation["building"](area.boundaryarea);
);
out center;
"""
```

```
# Send the query to the Overpass API to get response
response = requests.get(url, params={'data': query})
```

```
# Response variable returns the response code like 200, 400, 404 etc. Ideally it should be 200.
response.json()
```

```
                      'tags': {'highway': 'footway', 'surface': 'brick'}},
                    {'type': 'way',
                     'id': 297971591,
                     'center': {'lat': 40.4858921, 'lon': -74.4370514},
                     'nodes': [3018710884,
                      3018710885,
                      3018710886,
                      3018710887,
                      3018710888,
                      3018710889,
                      3018710890,
                      3018710891,
                      3018710892,
                      9002327436,
                      4425308127],
                     'tags': {'bicycle': 'yes',
                      'highway': 'footway',
                      'incline': 'up',
                      'name': 'Sacred Path',
                      'surface': 'brick'}},
                    {'type': 'way',
                     'id': 297971592,
                     'center': {'lat': 40.4868402, 'lon': -74.4373561},
                     'nodes': [3018620943, 3018710894, 3018710896, 3018710895],
                     'tags': {'highway': 'path'}},
                    {'type': 'way',
                     'id': 297971593,
                     'center': {'lat': 40.4866775, 'lon': -74.437748},
                     'nodes': [3018710896,
                      3018710897,
                      3018710903,
                      3018710898,
                      3018710899,
                      3018710900,
                      3018710901,
                      3018620937],
                     'tags': {'highway': 'path'}},
                    {'type': 'way',
                     'id': 297971595,
                     'center': {'lat': 40.4867271, 'lon': -74.4371288},
                     'nodes': [3018710894, 3018710904, 3018710929],
                     'tags': {'highway': 'path'}},
                    {'type': 'way',
                     'id': 297971596,
                     'center': {'lat': 40.4865092, 'lon': -74.4370311},
                     'nodes': [3018710905, 3018710906, 3018710907, 3018718667, 3018710908],
                     'tags': {'highway': 'path'}},
                    ...]}
```

▼ (Que 3) To get data in geojson, out body geom and [out:json] is passed in params. For text, .text method is used.

```
# Query params to get output in geojson
geojson_query = """
[out:json];
area["name"="New Brunswick"]->.boundaryarea;
(
  way["highway"](area.boundaryarea);
  relation["highway"](area.boundaryarea);
  way["building"](area.boundaryarea);
  relation["building"](area.boundaryarea);
);
out body geom;
"""


# Send the query to the Overpass API
geojson_response = requests.get(url, params={'data': geojson_query})


# To convert the response to text. This can be used with json or XML response too.
geojson_response.text
```

```
    '{\n  "version": 0.6,\n  "generator": "Overpass API 0.7.59.2 0994154d",\n  "osm3s": {\n    "timestamp_osm_base": "2023-02-23
    T13:00:27Z",\n    "timestamp_areas_base": "2023-02-23T10:44:06Z",\n    "copyright": "The data included in this document is f
    rom www.openstreetmap.org. The data is made available under ODbL."\n  },\n  "elements": [\n\n{\n  "type": "way",\n  "id": 11
    648485,\n  "bounds": {\n    "minlat": 40.5053361,\n    "minlon": -74.4518320,\n    "maxlat": 40.5056240,\n    "maxlon": -74.
    4515126\n  },\n  "nodes": [\n    4394181032,\n    6910757547,\n    104131645\n  ],\n  "geometry": [\n    { "lat": 40.505336
    1, "lon": -74.4515126 },\n    { "lat": 40.5055309, "lon": -74.4517482 },\n    { "lat": 40.5056240, "lon": -74.4518320 }\n
    ],\n  "tags": {\n    "HFCS": "Urban Minor Arterial",\n    "NJDOT_SRI": "12000672",\n    "cycleway:right": "lane",\n    "high
    way": "tertiary",\n    "lanes": "1",\n    "oneway": "yes",\n    "ref": "(672)",\n    "tiger:cfcc": "A41",\n    "tiger:count
```

Writing the outputs to different file formats.

▾ (Que 4) Getting filtered data.

> We have filtered data using:
>
> - City - Fredericton
> - Street Name - King Street, Regent Street etc.
> - Tags - Highway, One Way etc.

```
# API query params for filtering on basis of city in NB
city_query = """
[out:json];
area["name"="Fredericton"]->.boundaryarea;
way(area.boundaryarea)["highway"];
out tags;
"""
```

```
# Send the request to the Overpass API and get the response
city_response = requests.get(url, params = {"data": city_query})
```

```
# Converts the city_response object to json format
city_response.json()
```

```
'lit': 'no',
'name': 'Rue Saint-Roch',
'sidewalk': 'none',
'surface': 'asphalt'}},
...]}
```

▼ Filtering data for a particular street in Freddy

```
# Street query params for Fredericton
street_query = """
[out:json];
area[name="Fredericton"]->.fd;
node(area.fd)["addr:street"="King Street"];
out;
"""

# Send the HTTP request to the Overpass API and retrieve the JSON response
street_response = requests.get(url, params={'data': street_query})

# Returns the response in json format.
street_response.json()
```

```
        'lon': -66.6455949,
        'tags': {'addr:city': 'Fredericton',
         'addr:housenumber': '334',
         'addr:postcode': 'E3B 1E3',
         'addr:street': 'King Street',
         'name': 'Victory Meat Market',
         'opening_hours': 'Mo-Su 09:00-09:00',
         'shop': 'supermarket'}},
       {'type': 'node',
        'id': 3717803095,
        'lat': 45.9620341,
        'lon': -66.6442607,
        'tags': {'addr:city': 'Fredericton',
         'addr:housenumber': '400',
         'addr:street': 'King Street',
         'name': 'Robert Simmons',
         'shop': 'clothes'}},
       {'type': 'node',
        'id': 6201975552,
        'lat': 45.9618024,
        'lon': -66.6429524,
        'tags': {'addr:city': 'Fredericton',
         'addr:housenumber': '441',
         'addr:postcode': 'E3B 1E5',
         'addr:street': 'King Street',
         'check_date': '2023-02-05',
         'name': 'Savage`s Bicycle Centre',
         'shop': 'bicycle'}},
       {'type': 'node',
        'id': 6515020478,
        'lat': 45.9606302,
        'lon': -66.6415038,
        'tags': {'addr:city': 'Fredericton',
         'addr:housenumber': '520',
         'addr:street': 'King Street',
         'name': 'The Happy Baker',
         'shop': 'bakery'}},
       {'type': 'node',
        'id': 7612655976,
        'lat': 45.961931,
        'lon': -66.6431848,
        'tags': {'addr:city': 'Fredericton',
         'addr:housenumber': '435',
         'addr:street': 'King Street',
         'check_date': '2023-02-05',
         'disused:amenity': 'marketplace',
         'name': 'The Cultural Market',
         'opening_hours': 'Mo-Fr 07:30-18:00; Su 07:30-14:00',
         'opening_hours:covid19': 'open'}},
       {'type': 'node',
        'id': 10282747768,
        'lat': 45.9623021,
        'lon': -66.6442157,
        'tags': {'addr:city': 'Fredericton',
         'addr:housenumber': '397',
         'addr:street': 'King Street',
         'name': 'Tip Top Barbershop',
         'shop': 'hairdresser'}}]]
```

▾ Filtering data for json data with all streets or roads having highway in their tags

```python
# Query params for getting highway details in Fredericton
highway_query = """
[out:json];
area[name="Fredericton"]->.nb;
way(area.nb)["highway"];
out tags;
"""

# Define the API parameters
params = {
    "data": highway_query,
    "out": "tags",
}


# Send the HTTP request to the Overpass API and retrieve the response
highway_response = requests.get(url, params=params)


# Returns the list in json format.
highway_response.json()
```

...⌡⌡

(Que 5) Streaming APIs: Collect Twitter or any other streaming API data for 1 minute on your favourite topic and save it in a
JSON file

```python
import praw
import time
import warnings


# Filter out the warning message on console
warnings.filterwarnings("ignore", category=UserWarning, module="praw")


# Create a Reddit instance
reddit = praw.Reddit(client_id='client ID', client_secret='Client Secret', user_agent='gge/1.0 (sonaljo13@gmail.com)',check_for_as

# Set the subreddit to fetch comments from
subreddit = reddit.subreddit('AskReddit')


# Create an empty list to store the comments
comments = []


# Defining the 1 minute time for live streaming
end_time = time.time() + 60

# Using the reddit praw methods to stream the live comments from AskReddit group
for comment in subreddit.stream.comments(skip_existing=True):
  comment_data = {
          'body': comment.body,
          'author': comment.author.name,
          'created_utc': comment.created_utc,
          'subreddit': comment.subreddit.display_name,
      }
  print(comment_data)
  comments.append(comment_data)
# Breaks the loop when the 60 seconds condition is met
  if time.time() >= end_time:
    break
```

```
{'body': 'Kpop idols', 'author': 'alien-toe-girl', 'created_utc': 1677157500.0, 'subreddit': 'AskReddit'}
{'body': "As far as sexual partners goes, not really. As far as murders goes, yes. Absolutely. I'd go as far as to say that t
{'body': '**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**\n\nHi there, your post has been removed for one
{'body': 'Yep, been friends for about 8 to 9 years now', 'author': 'jayniir', 'created_utc': 1677157501.0, 'subreddit': 'AskR
{'body': 'Be able to relive past memories as if youve travelled back in time', 'author': 'Plastic_Afternoon_81', 'created_utc
{'body': 'Bees', 'author': 'Super-Noodles', 'created_utc': 1677157506.0, 'subreddit': 'AskReddit'}
{'body': 'Shout out to all of us with imposter syndrome!', 'author': 'sleep_envy', 'created_utc': 1677157506.0, 'subreddit':
{'body': 'at this point\n\nthings have gone to such shit\n\nits either laugh\n\n or go drink cyanide and play in traffic', 'a
{'body': 'AIDS', 'author': '_notyalc_', 'created_utc': 1677157507.0, 'subreddit': 'AskReddit'}
{'body': 'Your post nearly made my eyes roll out of my head.', 'author': 'cacotopic', 'created_utc': 1677157507.0, 'subreddit
{'body': "I love our mechanic. Little local shop, fair prices, nice folks, and only a few blocks away so I can just walk to p
{'body': "That would be in the US, but in countries where insurances don't control the healthcae market, this could be seen."
{'body': 'If its a good sleepover then gossip, snacks, movies. if it a great sleepover, witchcraft.', 'author': 'StygianBlueS
{'body': 'Fucking my cat', 'author': 'Character-Ad-4546', 'created_utc': 1677157509.0, 'subreddit': 'AskReddit'}
{'body': 'That vaccines can give you autism', 'author': 'iStoleYourSoda', 'created_utc': 1677157510.0, 'subreddit': 'AskReddi
{'body': 'That your facial hair grows thicker the more you shave', 'author': 'BbylonPrince', 'created_utc': 1677157510.0, 'su
{'body': "Don't eat right before bed", 'author': 'SlimThickChick3', 'created_utc': 1677157512.0, 'subreddit': 'AskReddit'}
{'body': 'I almost threw up', 'author': 'PixalPop', 'created_utc': 1677157513.0, 'subreddit': 'AskReddit'}
{'body': 'Tinder', 'author': 'LesseZTwoPointO', 'created_utc': 1677157516.0, 'subreddit': 'AskReddit'}
{'body': "My dog licks the mud and duck shit off my barn boots twice a day. I can't get him to stop. Send help", 'author': 'I
{'body': 'The need to pee\n\nA career I enjoy\n\nUp to 11 paid weeks off each year/multiple vacations', 'author': 'Supercheck
{'body': 'That 30 is old. I'm 26 now and for the first time in my life i know who i really am what i enjoy. My 30s will be th
{'body': "\n\ni'm from brisbane and i love it! it's a great city with a lot of culture, amazing food and drink, and plenty of
{'body': 'Clean towels, hot out of the dryer', 'author': 'Few-Environment-1597', 'created_utc': 1677157519.0, 'subreddit': 'A
{'body': 'It actually makes you pretty shitty person', 'author': 'manaha81', 'created_utc': 1677157519.0, 'subreddit': 'AskRe
{'body': '> I think I may be paying for your plan.\n\nAs you should be.  What do you think health insurance is?\n\nThe money
{'body': 'Being overweight doesn't make you less important/valued. Took me about 23 years to realize that.', 'author': 'broad
{'body': 'He's saying OSHA is crippling while at the same time not actually succeeding in its objective. As someone aptly men
{'body': 'Make all drugs legal and let them kill themselves off', 'author': 'NoiseRemarkable5012', 'created_utc': 1677157522.
{'body': 'Naval radar was also key in BVR targeting. That's why the giants Yamato and Musashi were fairly useless despite fie
{'body': 'Onlyfans', 'author': 'HateUsCuzDeyAunus', 'created_utc': 1677157524.0, 'subreddit': 'AskReddit'}
{'body': 'Teolological argument', 'author': 'Key_Veterinarian3191', 'created_utc': 1677157525.0, 'subreddit': 'AskReddit'}
{'body': 'anal seepage', 'author': 'Ahshalon_Tenisk', 'created_utc': 1677157525.0, 'subreddit': 'AskReddit'}
```

```
{'body': 'I was maybe ten. we all slept outside on this girl's trampoline. Someone put shaving cream in every girl's hair exc
{'body': 'All the anti science stuff.', 'author': 'SisterYahtzee', 'created_utc': 1677157529.0, 'subreddit': 'AskReddit'}
{'body': "Wired headphones don't need charging, have better quality, don't disconnect, are cheaper, are harder to lose, and d
{'body': 'Can't come because I'm sick', 'author': 'Chocolatelover4ever', 'created_utc': 1677157531.0, 'subreddit': 'AskReddi
{'body': "That you should be loyal to a company or business or manager. Your loyalty to a business is to do your job that you
{'body': 'Moonstruck, The Bird Cage', 'author': 'steel_city_sweetie', 'created_utc': 1677157533.0, 'subreddit': 'AskReddit'}
{'body': 'Either they stick with the actor or they recast.', 'author': 'CommodoreKrusty', 'created_utc': 1677157534.0, 'subre
{'body': 'At the time of replying, everyone is misinterpreting this as you being pro-speedtrap. Just wanna say that I underst
{'body': "No bother. My wife and I have created probably half a dozen including us and other people. My only concern is that
{'body': 'And the fact we are debatably the cleanest motherfuckers around.', 'author': 'UngusBungus_', 'created_utc': 167715
{'body': "Isn't it though?", 'author': 'See_Bee10', 'created_utc': 1677157536.0, 'subreddit': 'AskReddit'}
{'body': 'Wasted no time with that one', 'author': 'Tbrock8888', 'created_utc': 1677157538.0, 'subreddit': 'AskReddit'}
{'body': 'I haven't super looked into this but it is a fact. It's usually not like.. just feeling someone watching you though
{'body': 'You motorboating son of a bitch', 'author': '_-__-__-__-__-_-__', 'created_utc': 1677157540.0, 'subreddit': 'AskF
{'body': 'Well, Medicare is by far the largest and most efficient health insurer, so there's that one. There's the forest se
{'body': "Money doesn't motivate me at all. It's normally my bladder that does it.", 'author': 'lizzpop2003', 'created_utc':
{'body': 'New Car Smell', 'author': 'KENKAIYT', 'created_utc': 1677157544.0, 'subreddit': 'AskReddit'}
{'body': 'if im up before my wife\n\ni get to watch her get ready for work', 'author': 'Ahshalon_Tenisk', 'created_utc': 167
{'body': 'Btw this all got here from me staring at American television right here in uk and I've noticed a few extraordinary
{'body': 'Oh, so by "unsuspecting job", you mean like "jobs hold by people who don\'t suspect shit"? Well then many replaceat
```

```python
# Prints the data in the variable comments
comments
```

```
    {'body': 'And the fact we are debatably the cleanest motherfuckers around.',
     'author': 'UngusBungus_',
     'created_utc': 1677157535.0,
     'subreddit': 'AskReddit'},
    {'body': "Isn't it though?",
     'author': 'See_Bee10',
     'created_utc': 1677157536.0,
     'subreddit': 'AskReddit'},
    {'body': 'Wasted no time with that one',
     'author': 'Tbrock8888',
     'created_utc': 1677157538.0,
     'subreddit': 'AskReddit'},
    {'body': 'I haven't super looked into this but it is a fact. It's usually not like.. just feeling someone watching you
    though, it's more of the scenario. Walking down a Dark road? Might get that feeling, even if no one is even near you. Just
    because there MIGHT be a predator. But for when there is a predator that's when it's usually pretty damn handy. Predator
    being snything from an animal to some creep',
     'author': 'CanlexGaming',
     'created_utc': 1677157538.0,
     'subreddit': 'AskReddit'},
    {'body': 'You motorboating son of a bitch',
     'author': '_-__-__-__-__-_-__',
     'created_utc': 1677157540.0,
     'subreddit': 'AskReddit'},
    {'body': 'Well, Medicare is by far the largest and most efficient health insurer, so there's that one. There's the forest
    service and the bureau of land management which oversees vast woodlands and natural resources. The state and national parks
    systems – haven't seen any corporate parks developed yet. Certainly none without an inducement from… the government. There's
    the social security administration which oversees the largest pension, for which few private comparisons even exist anymore.
    You've got the federal bankruptcy court system which is a model for the rest of the world and is left to clean up the pieces
    when corporations fail, so let's count that one. \n\nSeems to me like you haven't spent much time around corporations.
    They're not infallible efficiency machines. They're capacity constrained just like everything else. Waiting in line at the
    DMV doesn't mean they're doing a bad job, it means that they have limited throughput and have to service everyone who comes
    calling in the order they arrived. \n\nI don't know about you, but I can't remember the last time someone "popped over" and
    asked to help as soon as I walked into a store. \n\nMaybe I'm just not weighting long enough for them to come by.',
     'author': 'Dismal_Operation_933',
     'created_utc': 1677157541.0,
     'subreddit': 'AskReddit'},
    {'body': "Money doesn't motivate me at all. It's normally my bladder that does it.",
     'author': 'lizzpop2003',
     'created_utc': 1677157542.0,
     'subreddit': 'AskReddit'},
    {'body': 'New Car Smell',
     'author': 'KENKAIYT',
     'created_utc': 1677157544.0,
     'subreddit': 'AskReddit'},
    {'body': 'if im up before my wife\n\ni get to watch her get ready for work',
     'author': 'Ahshalon_Tenisk',
     'created_utc': 1677157547.0,
     'subreddit': 'AskReddit'},
    {'body': 'Btw this all got here from me staring at American television right here in uk and I've noticed a few
    extraordinary provides already',
     'author': 'Ecstatic_Garbage_571',
     'created_utc': 1677157549.0,
     'subreddit': 'AskReddit'},
    {'body': 'Oh, so by "unsuspecting job", you mean like "jobs hold by people who don\'t suspect shit"? Well then many
    replaceable jobs, even with today\'s tech, are unsuspecting lol.',
     'author': 'ZigotoDu57',
     'created_utc': 1677157554.0,
     'subreddit': 'AskReddit'}]
```

```
# Save the comments to a JSON file
with open('comments.json', 'w') as f:
    f.write(json.dumps(comments))
```

✓ 0s   completed at 9:06 AM   ● ✕