

Overall Summary

paul

March 30, 2017

Goal:

The goal is to look into how to access APIs for various data sources: Facebook, twitter, Instagram/Flickr and weather purely through the usage of R.

Facebook:

- **Advantages:**

- Easy to retrieve data from facebook API using the functions provided by the **Rfacebook**.
 - Easy to access public data of facebook through its **search API**. Part of the functionality is covered by the *Rfacebook* package through the *searchGroup* & *searchPages* function.
 - Long lived tokens created through registering an facebook application does not expire.
 - Discussions about a post can be easily tracked through extracting comments and replies to those comments.
 - The graph API Explorer tool is useful for testing API calls to understand structure as well as exploring possible fields and edges.
 - The **Utils.R** package within *Rfacebook* provides example frameworks to parse the response JSON nested list into an organized data frame.
 - To use custom API requests, use the *CallAPI* function in *Rfacebook* with the request URL to retrieve response.
-

- **Disadvantages:**

- The facebook API documentation does not provide specific information regarding **Rate Limit**, it specifies that the limit **differs depending on request**.
 - facebook API V2.0 and above place heavy restrictions on accessing personal information. Details such as profile information, user timeline etc. are not generally publicly available due to privacy settings. It is **not recommended** to scrape the facebook API to search for user related information.
 - To retrieve information regarding nodes, the Id of the node is required. This makes tracking objects/searching for specific nodes very difficult since there is no easy way to obtain the id for specific objects of interest (pages, users, comments, etc.)
 - It was observed that many pages, posts and comments does not include *location* information as part of its content.
-

Facebook Detailed Report

Facebook Report

Twitter:

- Twitter API
-

- **Advantages:**

- The packages provide easy access to the API.
 - **It is recommended to retrieve *RECENT* information using the Twitter API compared to Facebook API.** This is because most information posted on twitter is public (tweets, users, etc).
 - The search API for Twitter supports more powerful queries for searching public twitter content.
 - The twitterR Package provide easy access to the Twitter REST API and handles the sorting of retrieved data through class wrappers that provide convenient functions to access information. Tweets are part of the *status* class for example. Read the documentation for useful details.
 - TwitterR package retrieves majority of information that is of interest, if not supported by the twitterR, any arbitrary HTTP request can be made to the API to retrieve information.
 - Twitter provides a live stream API which allows the retrieval of the new updated information on tweeter. This is a unique feature and can be used for real-time tracking for events of interest.
 - StreamR provides easy access and convenient use of the stream API connecting, saving, parsing of data is all handled.
-

- **Disadvantages:**

- Location information is available but often users can fill whatever it desired, thus not most accurate.
 - The token produced by twitterR is cached internally and not provide easy access to the token for use outside of twitterR functions.
 - Due to nature of twitter, it is hard to track how users react to tweets posted by other users. This is because the replies is not directly connected to the post.
 - **It is not possible to public search for old tweets (older than a week) through the official twitter API.**
-

Twitter Report:

Twitter Report

Google Maps API:

Google Maps API

- **Advantages:**

- Provides many services including directions between origin and destination, geocoding services, matching points to near roads, elevation, etc.
- Documentation is clear and very easily understood.
- Use basic **HTTP request** with simple format, making it easy to use.

- **Disadvantages:**

- limited free requests per day. The free limit is enough for investigation purposes. If used to process large amounts of data, a fee will need to be paid.
-

Google Maps Report:

Google Maps Report

Quick Word on OpenStreetMap(OSM):

Open Street Map Wiki

Preliminary Report on OSM

Weather Web Data:

An basic script was made to scrape the Environment Canada Ottawa website for hourly updated temperature and 24 hour forecast. The collected data was then explored by making several graphs.

- Weather Web Scraping
 - Data Exploration
-

Instagram:

- From initial searching, the Instagram API is not publicly available. In order to access the API, an Instagram approved application (only applications that promote sharing through Instagram can be approved). Since the API is not available, not further search into Instagram was done. Instagram Permissions Review.
-

Image Metadata:

- As part of the search into processing images through Instagram, a command prompt tool for windows called exifTool, which allows the extraction of images metadata.
 - Image Metadata Brief Report
-

Flickr:

- Flickr is a photo sharing website which allows people to host their photos online.
 - Flickr API
 - **The API is not maintained or supported by Flickr.**
-

- **Advantages:**
 - Allows public access with relative ease.
 - Many methods does not require OAuth 1.0 token.

- Photos usually contain meta data.
 - Photo tags provide an easy way to adjust search query.
 - **Disadvantages:**
 - Does not have massive amounts of data available like twitter. This is because many photos are not public.
 - Limited requests.
-

Flickr Report