

Scraping Reddit using Python: instructions:

1. Creating a Reddit App.

The first thing you need to do to get access to the reddit API data is to create a reddit app. You can do that by going to this link: <https://www.reddit.com/prefs/apps>

After you log in with your reddit account you will be able to see a button that says "Are you a developer? create an app..".

When you press the button you should see something like this:

create application

Please [read the API usage guidelines](#) before creating your application. After creating, you will be required to [register](#) for production API use.

name

bds516_scraper

web app

installed app

script

A web based application

An app intended for installation, such as on a mobile phone

Script for personal use. Will only have access to the developers accounts

description

The app to get data for reddit for the BDS516 project

about url

redirect uri

http://localhost:8080

create app

You need to do the following:

- Choose the name for the app (it can be anything, but make sure it's one word)
- Make sure you choose "script" for the type of the app.
- You can add a description
- The redirect uri should be: <http://localhost:8080>

Once the app is created, you will be able to seee three things:

- client_id (right under the words personal use script)
- secret (right night to the words secret
- user_agent (that's the name you gave to your app)

1. Set up your python script.

Install the package "praw" (it can be done once per lifetime of a computer)

pip install praw

```
In [2]: !pip install praw # comment this line out if you installed it before.

Collecting praw
  Downloading praw-7.5.0-py3-none-any.whl (176 kB)
    |████████████████████| 176 kB 12.5 MB/s eta 0:00:01
Requirement already satisfied: websocket-client>=0.54.0 in /Users/shpenev/opt/anaconda3/lib/python3.8/site-packages (from praw) (0.58.0)
Collecting prawcore<3,>=2.1
  Downloading prawcore-2.3.0-py3-none-any.whl (16 kB)
Collecting update-checker>=0.18
  Downloading update_checker-0.18.0-py3-none-any.whl (7.0 kB)
Requirement already satisfied: requests<3.0,>=2.6.0 in /Users/shpenev/opt/anaconda3/lib/python3.8/site-packages (from prawcore<3,>=2.1->praw) (2.27.1)
Requirement already satisfied: certifi>=2017.4.17 in /Users/shpenev/opt/anaconda3/lib/python3.8/site-packages (from requests<3.0,>=2.6.0->prawcore<3,>=2.1->praw) (2021.10.8)
Requirement already satisfied: idna<4,>=2.5 in /Users/shpenev/opt/anaconda3/lib/python3.8/site-packages (from requests<3.0,>=2.6.0->prawcore<3,>=2.1->praw) (3.3)
Requirement already satisfied: charset-normalizer~=2.0.0 in /Users/shpenev/opt/anaconda3/lib/python3.8/site-packages (from requests<3.0,>=2.6.0->prawcore<3,>=2.1->praw) (2.0.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /Users/shpenev/opt/anaconda3/lib/python3.8/site-packages (from requests<3.0,>=2.6.0->prawcore<3,>=2.1->praw) (1.26.8)
Requirement already satisfied: six in /Users/shpenev/opt/anaconda3/lib/python3.8/site-packages (from websocket-client>=0.54.0->praw) (1.16.0)
Installing collected packages: update-checker, prawcore, praw
Successfully installed praw-7.5.0 prawcore-2.3.0 update-checker-0.18.0
```

Now load the package and set up the reddit instance. Replace the info with your own data

```
In [ ]: import praw

reddit = praw.Reddit(client_id='TYPE YOUR CLIENT ID HERE',
                     client_secret='TYPE YOUR CLIENT SECRET HERE',
                     user_agent='TYPE YOUR USER AGENT HERE')
```

Now you can use the API to scrape the data from Reddit.

Suppose you want to see the 10 most popular posts in the subreddit "Cats"

```
In [5]: import pandas as pd
posts = []
subreddit = reddit.subreddit('Cats')
for post in subreddit.hot(limit=10):
    posts.append([post.title, post.score, post.id, post.subreddit, post.url, post.num_comments, post.selftext, post.created])
posts = pd.DataFrame(posts,columns=['title', 'score', 'id', 'subreddit', 'url', 'num_comments', 'body', 'created'])
print(posts)

      title  score  id subreddit \
0      Chilling in the shower 10341  ultsh7      cats
1  Video games are great, but sometimes it's nice... 2007  ulypu9      cats
2  He's trying to earn his keep. He's gone to ten... 1521  ulz8nd      cats
3      What kind of cat do I have? 1077  u21p6l      cats
4  My boyfriend died two weeks ago and my cat has... 1094  ulziv2      cats
5  Spidercat, spidercat, does whatever a spiderca... 759   u20x8j      cats
6              just chillin 551   u22jf8      cats
7              cutie beau and cool 8107  ulo69k      cats
8      My cat Bucket on his homemade couch 563   u2012d      cats
9      Catto Wanna Hide 7158   ulnrvf      cats

      url  num_comments  body  created
0  https://i.redd.it/yoleio2ul2t81.jpg 120  1.649751e+09
1  https://v.redd.it/qbrlgstek3t81 22  1.649769e+09
2  https://i.redd.it/greqxhpuo3t81.jpg 167  1.649770e+09
3  https://i.redd.it/zyft6rnz84t81.jpg 257  1.649777e+09
4  https://www.reddit.com/gallery/ulziv2 50  1.649771e+09
5  https://i.redd.it/vel5za6y24t81.jpg 7  1.649775e+09
6  https://i.redd.it/fho4zxumf4t81.jpg 43  1.649779e+09
7  https://www.reddit.com/gallery/ulo69k 121  1.649730e+09
8  https://www.reddit.com/gallery/u2012d 22  1.649773e+09
9  https://v.redd.it/olvwdqld80t81 42  1.649729e+09
```

You can also get comments to posts using the API

```
In [7]: submission = reddit.submission(url="https://www.reddit.com/r/cats/comments/u21dcf/elder_stray_needs_to_be_adopted_he_roams_around/")

submission.comments.replace_more(limit=0)
for comment in submission.comments.list():
    print(comment.body)
```

Hello, I know some friends and family members who live in India and can possibly help, please can you message me privately more info I wish I could adopt him ...poor baby
I wish I could adopt him ❤️.But, I can't, I already have a few fur babies.
Wishing you and him the best of luck. Please find him a good him!
Thanks you 🙏
Done. I've DMed ya.
Thanks a lot. I wish I could too. My allergies unfortunately don't afford me the luxury. It was manageable before Covid, but it's been a mess ever since.

You can get inspiration for some additional commands here:

<https://medium.com/analytics-vidhya/scraping-reddit-using-python-reddit-api-wrapper-praw-5c275e34a8f4> <https://www.geeksforgeeks.org/scraping-reddit-using-python/>
<https://gilberttanner.com/blog/scraping-redditdata>