Name: Sayyed Sohail Rashid	Course Name: SMA-LAB
Class: BE-CO	Batch: 01
Roll no: 18CO48	Experiment No: 02

Aim: Data Collection-Select the social media platforms of your choice(Twitter, Facebook, LinkedIn, YouTube, Web blogs etc) ,connect to and capture social media data for business (scraping, crawling, parsing).

Theory:

Social Media Scraping of Apple using Python:

1. Instagram:

```
2. from instagramy import InstagramUser
3. import pandas as pd
4.
5. user = InstagramUser("Apple")
6. print(f"Username: {user.fullname}")
7. print(f"Biography: {user.biography}")
8. print(f"Verified User: {user.is_verified}")
9. print(f"Website: {user.website}")
10.print(f"Followers: {user.number_of_followers}")
11.print(f"Following: {user.number_of_followings}")
12.print(f"No. Of Posts: {user.number_of_posts}")
13.
14.posts = user.posts
15.instaPosts = []
16. for i in range(10):
17.     post = {}
18.     post["Post"] = i+1
19.     post["Likes"] = posts[i].likes
20.     post["Comments"] = posts[i].comments
```

```
21.     post["Caption"] = posts[i].caption
22.     instaPosts.append(post)
23. print(f"Latest posts:\n {pd.DataFrame(instaPosts)}")
24.
```

Output

```
Username: APPLE
Biography: Everyone has a story to tell.
 """#ShotoniPhone to take part.
Visit Apple Studio for more fashion & style advice.
Verified User: True
Website: https://apple.onelink.me/eWcy/gnczbjm9
Followers: 2947991
Following: 9
No. Of Posts: 1093
Latest posts:
Post         Likes  Comments Caption
0       1   3006        62     None
1       2   2626        43     None
2       3  14290        46     None
3       4    577        14     None
4       5    793         8     None
5       6    742        20     None
6       7   1242        55     None
7       8   1020        23     None
8       9   2043        16     None
9      10    349        27     None
```

2) Facebook:

```
25. # Facebook
26. from facebook_scraper import get_posts
27.
28. fbPosts = []
29. for post in get_posts('Apple', pages=10):
30.     fbPosts.append({x: post[x] for x in ['post_id', 'text', 'likes']})
31. print(print(f"Latest posts:\n {pd.DataFrame(fbPosts)}"))
```

Output

```
Latest posts:
              post_id  text                                                      likes
    0    10161168767013221  YAAASSS, we can't believe our eyes either             20.0
    1    10161168690158221  YAAASSS, we can't believe our eyes eitherv%           .10.0
    2    10161167893718221  This Pink monochrome look is      2\n\n    @simran...  24.0
    3    10161166649103221  Whether you're going on a date, or giving your...     79.0
    4    10161165440628221  Mark your calendar & set your reminder 'cause ...    23.0
    5    10161162761178221  Jump into an unparalleled listening experience...    28.0
    6    10161164921723221  Shhhh...something extraordinary coming your way,...   69.0
    7    10161164580113221  Shhhh...something extraordinary coming your way,...   24.0
    8    10161164466943221  Shhhh...something extraordinary coming your way,...   20.0
    9    10161164422793221  Shhhh...something extraordinary coming your way,...   27.0
    10   10161163521578221  Denim game strong with pantaloons!\nCheck out ...    34.0
    11   10161163439843221  A race before the actual race!\nThe Early Bird...    28.0
```

3) Twitter:

```python
from twitter_scraper_selenium import get_profile_details

twitter_username = "Apple"
filename = "twitter_api_data"
get_profile_details(twitter_username=twitter_username, filename=filename)

import json

with open('twitter_api_data.json', 'r') as json_file:
    json_object = json.load(json_file)

print(json.dumps(json_object, indent=1))
```

2023-02-06 14:15:46,817 - twitter scraper selenium.profile details - INFO - Data Successfully Saved to C:\IBM-0\IBM\IPT\twitter_api_data.json

 "id": 89443197,
 "id_str": "89443197",
 "name": "Apple",
 "screen_name": "apple",|
 "location": "Bengaluru, India",
 "profile_location": null,
 "description": "If you never try, you\u201911 never know \ud83e\udd0d Shop NOW\ud83d\udc47\ud83c\udffb",
 "url": "https://t.co/wIDM4c18J",
 "entities": {
  "url": {
   "urls":

     "url": "https://t.co/wIDM4c18J",
     "expanded_url": "https://apple.onelink.me/ewcy/gnozbjm9",
     "display_url": "apple.onelink.me/eWcy/gnozbjm9",
     "indices": [
      0,
      23

    ],
  "description":
   "urls": []

1
 "protected": false,
 "followers_count": 692685,
 "fast_followers_count": 0,
 normal_followers_count": 692685,
 "friends_count": 37, "listed_count": 497,
 created_at": "Thu Nov 12 13:16:01 +0000 2009",
 "favourites_count": 5803,
 "utc_offset": null,
 "time_zone": null,
 geo_enabled": true,
 "verified": true,
 "statuses_count": 497237,
 "media_count": 8500,
 "lang": null,
 "status":

"lang": null,
"status": {
 created_at": "Mon Feb 06 08:43:35 +0000 2023",
 "id": 1622516340518162433,
 "id_str": "1622516340518162433",
 "text": "jack I hear you & I am here for you, Jeet! Rest assured, I have got this highlighted, and one of our cas\u2026 https://t.c o/069kk07W11",
 "truncated": true,
 "entities": {
  "hashtags": [],
  symbols": [],
  user_mentions": [

    "screen_name": "jack", "name":
    "jack",
    "id": 1600448314482118656,
    "id_str": "1600448314482118656",
    "indices": [
     0,
     15

   }
  ],
  "urls": [
   {
    "url": "https ://t.co/069kk07W11",
    "expanded_url": "https://twitter.com/i/web/status/1622516340518162433",
    "display_url": "twitter.com/i/web/status/1\u2026",
    "indices": [
     121,
     144

   }
         "Sprinklr", to status id": 1622515741483495424,
         to_status_id_str": "1622515741483495424",
         to user_ia": 1600448314482118656,
 "source": to user id str": "1600448314482118656",
 "in_reply_
 "in_reply_
 "in_reply_
 "in_reply_

Conclusion:

The data collection has been done for Apple from multiple websites such as Twitter, Facebook, Instagram. In order to connect and capture social mediadata for businesses (scraping, crawling, parsing).