

Email: ali.ameli@mail.um.ac.ir

LinkedIn: http://www.linkedin.com/in/aliameli

**B.SC. IN COMPUTER ENGINEERING, FERDOWSI UNIVERSITY OF MASHHAD** 

Documentation for extracting top 10 newest questions which was voted more than others through the last week for android:

### Solution for this specific problem:

- 1. Using Stack overflow **API** and get the Questions between two dates and filter them according to your needs.
- 2. Using **Selenium Web driver** to surf the "https://api.stackexchange.com/docs/questions" and filter the inputs according to your needs and get the result as a JSON file.
- 3. Implement a web Crawler with Beautiful soup library in python.

I chose to use **stack overflow API** because it is so convenient and fast. I also implemented a program with Selenium web driver to get the JSON file from the stack overflow and then give it to another program to create the web page which is needed. Two other solutions have some drawbacks:

- Selenium web driver is not fast enough to do such an assignment and also has a heavy overhead (because it should open a web browser to search the URL and enter some values and search the result)
- Implementing a web crawler for this problem cannot be a good and wise decision because it would be **blocked** by the website after a few requests (I have the same experience of crawling LinkedIn, Instagram and ninisite comments)

This website provided a complete documentation of stack overflow API:

https://api.stackexchange.com/docs

## Algorithm:

// at the first step, I get the current date of the system and then calculate the date of 7 days ago because we need the latest week

```
N = 7
date N days ago = datetime.now () - timedelta(days=N)
```

// at the second step, I call stack API and then fetch a query to the website. I want top 10 questions in the last week which has the "android" tag and sort them by their voting and order them desc

```
SITE = StackAPI('stackoverflow')
questions = SITE.fetch('questions', fromdate=date_N_days_ago, todate=datetime.now(),
pagesize=10, order='desc', tagged='android',sort='votes')
```

// at the third step, I call stack API and then fetch a query to the website. I want top 10 questions in the last week which has the "android" tag and sort them by their date

```
SITE = StackAPI('stackoverflow')
questions = SITE.fetch('questions', fromdate=date_N_days_ago, todate=datetime.now(),
pagesize=10, order='desc', tagged='android',sort='creation')
```

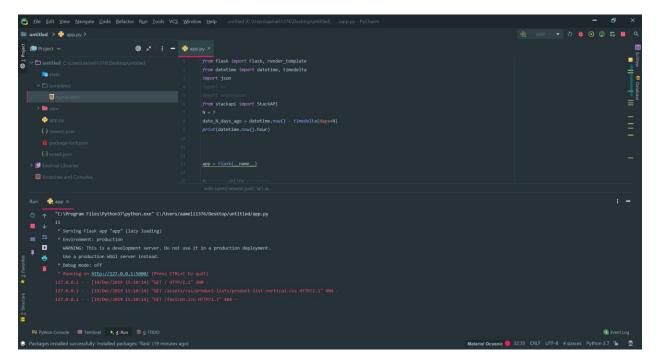
//then I save "questions" as a json file and then I processed this json file. I processed the JSON file with Flask in python to render t as HTML web page

#### How to run this code:

You just need to run the app.py file. It has some library that you must install them such as webbrowser and stackapi and Flask.

pip install Flask
pip install StackAPI

My python version is 3.7



#### **Results:**

When I run my code at first, I receive such a JSON file from the Stack Overflow website which consists of important information about each question:

```
1 ["backoff": 0, "has_more": true, "page": 5, "quota_max": 300, "quota_remaining": 300,
     "total": 0, "items": [{"tags": ["android", "admob", "google-play-services"], "owner": {"reputation": 351, "user_id": 4361781, "user_type": "registered", "profile_image":
     "https://i.stack.imgur.com/TltNg.jpg?s=128&g=1", "display_name": "Meet", "link":
     "https://stackoverflow.com/users/4361781/meet"}, "is_answered": true, "view_count": 303,
     "accepted_answer_id": 59283774, "answer_count": 2, "score": 6, "last_activity_date":
     1576131289, "creation_date": 1576057905, "last_edit_date": 1576131289, "question_id":
     59282970, "link":
     "https://stackoverflow.com/questions/59282970/play-install-referrer-library", "title":
     "Play Install Referrer Library"}, {"tags": ["android", "kotlin"], "owner": {"reputation":
     511, "user_id": 7750432, "user_type": "registered", "profile_image":
     "https://i.stack.imgur.com/yb97E.jpg?s=128&g=1", "display_name": "Yaroslav Ovdiienko",
     "link": "https://stackoverflow.com/users/7750432/yaroslav-ovdiienko"}, "is_answered": true,
      "view_count": 66, "accepted_answer_id": 59304966, "answer_count": 1, "score": 5,
     "last_activity_date": 1576154880, "creation_date": 1576070274, "last_edit_date":
     1576074016, "question_id": 59286707, "link":
     "https://stackoverflow.com/questions/59286707/hyperion-android-serviceloader-can-t-load-cus
     tom-plugin", "title": "Hyperion-Android: ServiceLoader can\u2019t load custom Plugin"},
     {"tags": ["android", "location", "android-permissions"], "owner": {"reputation": 79,
      'user_id": 11221923, "user_type": "registered", "profile_image":
     "https://lh3.googleusercontent.
     sz=128", "display_name": "ali mazaheri", "link":
     "https://stackoverflow.com/users/11221923/ali-mazaheri"}, "is_answered": true,
     "view_count": 55, "accepted_answer_id": 59285288, "answer_count": 1, "score": 4,
     "last_activity_date": 1576068290, "creation_date": 1576061719, "last_edit_date":
     1576068290, "question_id": 59284230, "link":
     "https://stackoverflow.com/questions/59284230/how-get-location", "title": "How get
     location?"}, {"tags": ["android", "kotlin", "textutils"], "owner": {"reputation": 355,
"user_id": 2897683, "user_type": "registered", "accept_rate": 93, "profile_image":
     "https://www.gravatar.com/avatar/5c27b9e59f49407f3c4e252df453d13b?s=128&d=identicon&r=PG&
     f=1", "display_name": "thahgr", "link": "https://stackoverflow.com/users/2897683/thahgr"},
     "is_answered": true, "view_count": 68, "accepted_answer_id": 59321544, "answer_count": 7,
     "score": 3, "last_activity_date": 1576244939, "creation_date": 1576235519, "question_id":
```

I saved this JSON file and process it. for example, I extract the score, link to the question, title of the question, number of answers to that question and number of views for that question.

This screenshot is the result which I printed in command line of PyCharm:

# **Functionality:**

The web page is **responsive** and **user-friendly**, I implemented it with Flask in python. By clicking on a title, the question content will be opened and you can see the question. By clicking on read more you can go to the source website and answer the question.

