

ALI AMELI

Iran, Mashhad · +98 937 882 5090

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

LinkedIn: <http://www.linkedin.com/in/aliamei>

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 it 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

3

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

Functionality:

[illegible]