

Midterm

Deadline : Nov 1st. 5:00 pm (1 hour before midterm)

Total Points : 100

Question

- Use the StackExchange REST API to collect the data for stackoverflow. You are supposed to use at least 3 per-site method. (40 points)
- API Documentation :
- [**https://api.stackexchange.com/docs**](https://api.stackexchange.com/docs)
- Perform 5 analysis on the data that you collect. (60 points)

Sample Analysis

- **Analysis 1 :**
- Get the Questions with tag python and pandas.
- Parse the body for the responses to collect a list of questions and user_id for the questions.
- Use the user_id obtained to send a request again to get user profile. Obtain the badges count to determine weightage.
- Show top questions according to the weightage.

Sample Analysis

- **Analysis 2 :**
- For each of the User ID that you have collected, ping the API to get all the tags that user has identified with.
- Create a file for each topic, containing user_id, user_name and link to their profile sorted by reputation.
- Question : For a given topic (say python), what are the top users who have reputation in that topic.

Sample Analysis

- **Analysis 3 :**
- For each of the badge type, find how many users (based on the data you have collected) have badge.
- Question : We want to see what badges are popular among the users.

Sample Analysis

- **Analysis 4 :**
- For each of the question that is asked, find out the tags attached to it.
- Find how many numbers of answers have been given for each question. For each tag, calculate the number of question asked and how many times it has been answered.

Sample Analysis

- **Analysis 5 :**
- Find out the user whose questions have been downvoted the most.