Project Bilibili web scraping

Bilibili (simplified Chinese: 哔哩哔哩; traditional Chinese: 嗶哩嗶哩), also nicknamed B Site (B站) in China, is a Chinese video sharing website based in Shanghai, themed around animation, comics, and games (ACG), where users can submit, view and add overlaid commentary on videos. Since the mid-2010s, Bilibili began to expand to a broader audience from its original niche market that focused on animation and games, and it has become one of the major Chinese video on demand over-the-top streaming platforms. Some people think that Bilibili will become a popular platform like Youtube in China.

Website link:

<https://www.bilibili.com/>

Please pay attention to the contract, and don’t do anything illegal.

<https://www.bilibili.com/robots.txt>

Requirement:

1. Get one user’s basic information by the user’s ID or by the name, including but not limited to his or her

Name, sex, rank, level, following and followers, Vip info. return one csv file with the above information.

1. Web scraping from the special zone to get the top 100 videos today or this month. For each video, get the Uper’ s name (Youtuber), videos title, released time, video id, time, comment numbers and so on. Return a csv file.
2. Try to download a user’s information in different formats, for example, videos, pictures (user’s face), text and others.
3. Mimic a user’s login process and give a thumb’s up or toss coins for all the videos of the users you like. (Or favorite his video in the past few months, a given period)
4. (Try to do a more challenging one) Use some Api in the internet and a more advanced crawler frame “Scrapy” to get a popular user’s basic information and focus on the user’s following and follower, and the follower’s followers. Then select the users who have more than 100 thousand fans to construct a small social network and save it in a database “SQL” or as a csv file.

Please note the following:

* Each proposal should include one website only.
* Please include the URL link and description of data to be scraped in the proposal.
* No need to include grading criteria and maximum completion time
* The proposals need to be approved before submission by the deadline in December.
* Points rewarded depend on complexity and originality. Quality is more important than quantity.