Project 5.31. JinRiTouTiao web scraping (a news website) and data storage

Introduction:

The code to scrap data from this website is relatively easy, where the only trick is the usage of json. However, the main point of this code is to do data collection in two ways, MongoDB and MySQL (Since using excel for data storage has been shown in my other projects).

Requirement:

1. Website link: <https://www.toutiao.com/>
2. Write a code to extract the following data:
   1. Title
   2. Image url
   3. url
3. Store the data by using MongoDB and MySQL respectively.
4. Bonus
   1. write PowerPoint slide to introduce the code (50/effective page)

Codes will be graded based on:

1. The format of the codes
2. The organization of the csv file
3. The speed of the codes
4. The extendibility of the codes

Estimated completion time: 25 hours. Time consumption depends on the codes’ quality.

Suggestive points:

Deadline:

Bonus points:

1. Work completion before the deadline will generate an extra 20% bonus.

Penalty:

1. Late delivery/missing deadline will incur -10% of finally determined points of the whole project per missing day.
2. Serious mistakes that made the submitted results not useful will incur -120% of the whole project.