**Project 5.26. Craig’s List Web Scraping**

Description of Project: Crawling the <http://hongkong.craigslist.hk> for flats/housing or holiday rentals. Such collected data can be used for market research in terms of what the real estate market is like in Hong Kong at a given time.

Link to Hong Kong’s craigslist flat/housing:

<https://hongkong.craigslist.hk/d/flats-housing-for-rent/search/apa>

Link to Hong Kong’s craigslist holiday rental:

<https://hongkong.craigslist.hk/d/holiday-rentals/search/vac>

Requirement: submit a code to realize the following function:

1. Enter the type of real estate data (i.e. flat/housing, holiday rental) you want to collect in the code;
2. Return a csv file with the following information:

A. Headline of Post

B. Date of Submission

C. Location of Real Estate

D. Number of Bedrooms

E. Size of Real Estate

F. Description within each post (e.g. <https://hongkong.craigslist.hk/vac/d/5-stars-airport-hotel-for/6402204926.html>)

a. post ID

b. last update

1. If entering a list of keywords, the code returns multiple csv files.

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: 70 hours. Time consumption depends on the codes’ quality.

Suggestive points:

Deadline:

Bonus points:

1.      Work completion one day 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.