**Project 5.11. Reddit Web Scraping**

Description of Project: Crawling the Reddit.com “WorldNews” subreddit for submissions about a specific company(s). Such collected data could then be used in research to e.g. investigate the correlation of media coverage and stock price (for example by feeding the headlines in a natural language processing network that determines the mood of the article headline (good news vs bad news, …) and thereafter predicts the stock price).

Link to Example Search (e.g. World News about Google): <https://www.reddit.com/r/worldnews/search?q=google&sort=new&restrict_sr=on&t=month>

Requirement: submit a code to realize the following function:

1. Enter the keyword that you want to search in the code;
2. Select if you want to crawl submitted headlines for the last month or last year
3. Return a csv file with the following information:

A. Headline

B. Time of Submission

C. Link to Source

D. Number of Comments

E. Number of Points (Upvote - for bonus points)

F. Name of Submitting User

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: 40 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.