

DSCI-558 HW1-Task 3

Student: Ziheng Gong

USC ID: 3429361173

3.1. What is the seed URL(s) you used?

- Task 1: `https://www.imdb.com/search/title/?genres=comedy`
- Task 2: `https://www.imdb.com/search/name/?gender=male,female`

3.2. How did you manage to only collect movie/show or cast pages?

For each task, the webpage from seed URL contains only the pages I need, and my parser also keeps only the pages I need.

3.3 Did you need to discard irrelevant pages? If so, how?

I did not need to discard irrelevant pages. Because all the `Request ()` objects my code generated are the pages I need.

3.4. Did you collect the required number of pages? If you were not able to do so, please describe and explain your issues.

I collected the required number of pages, about 5100 pages per task.