This is Baolu. My main job is data design and data collection. I collect and crawl data from web pages, clean up dirty data, and use sql to store in the database mysql. Designed the relationships and variables between tables to help the team link the database in the django framework and solve the data related questions of the team members (such as what data they need to retrieve, what comments to add, and what data format they want to use).

Thanks for your last time code submission feedback:

	10 pts Full Marks	6 pts Partial	0 pts No Marks	
WRITTEN: Specificity view longer description	**Comments** Appreciate the context to your submission! Good level of detail, I would also be interested in the tools you used to crawl and clean the data			10 / 10 pts

	5 pts Full Marks	3 pts Partial	0 pts No Marks	
DB: Diagrams				5 / 5 pts

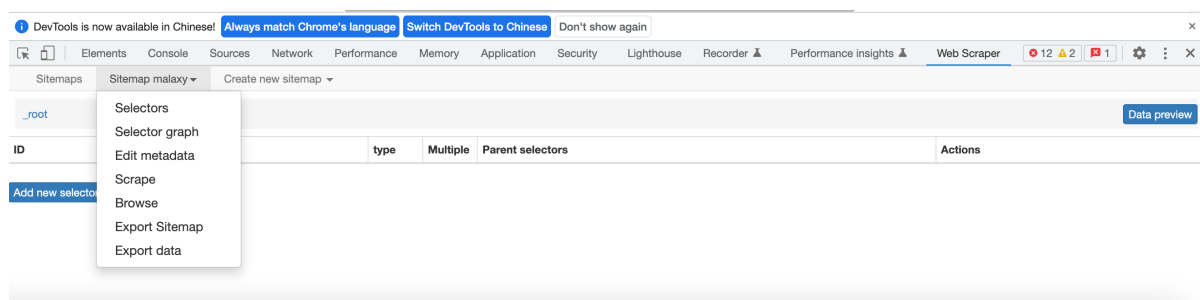|	5 pts Full Marks	3 pts Partial	0 pts No Marks	
DB: Documentation view longer description	**Comments** Good naming schema, SQL code			5 / 5 pts

Below is more things I did after last code submission:

1. We would like to make user experience more smoothly. To avoid using VPN to login to database everytime(since we used ischool's free server), we migrate some to SQLite, and after test, it works. You could also see my work in db.sqlite3.
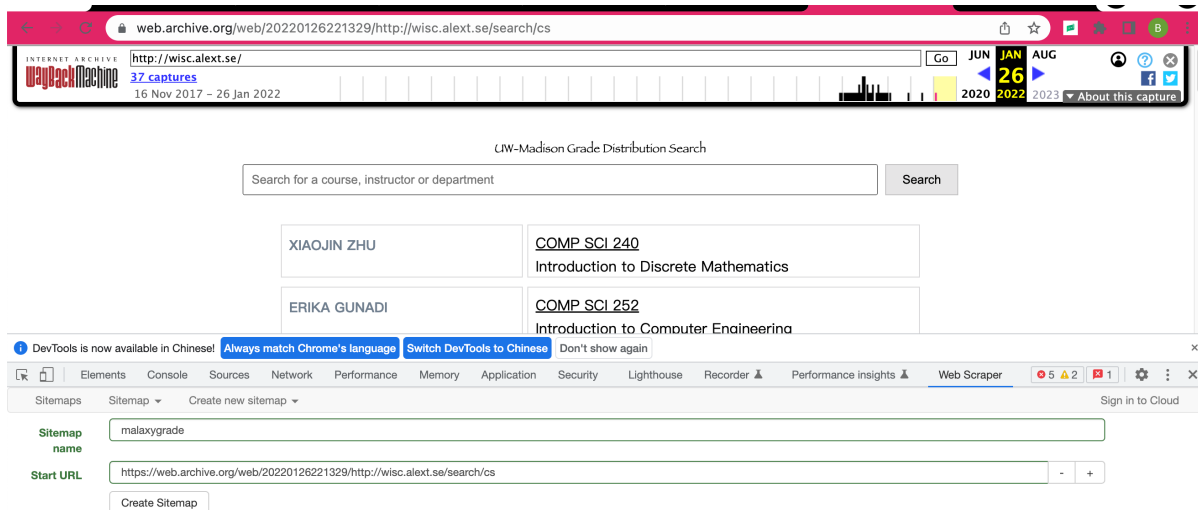
2. I updated more data to our project.(Section, Grade and Course tables)

Plus scraping tool I recommended is "WaybackMachine". (I saw the feedback from our last time code submission that you are interested in):
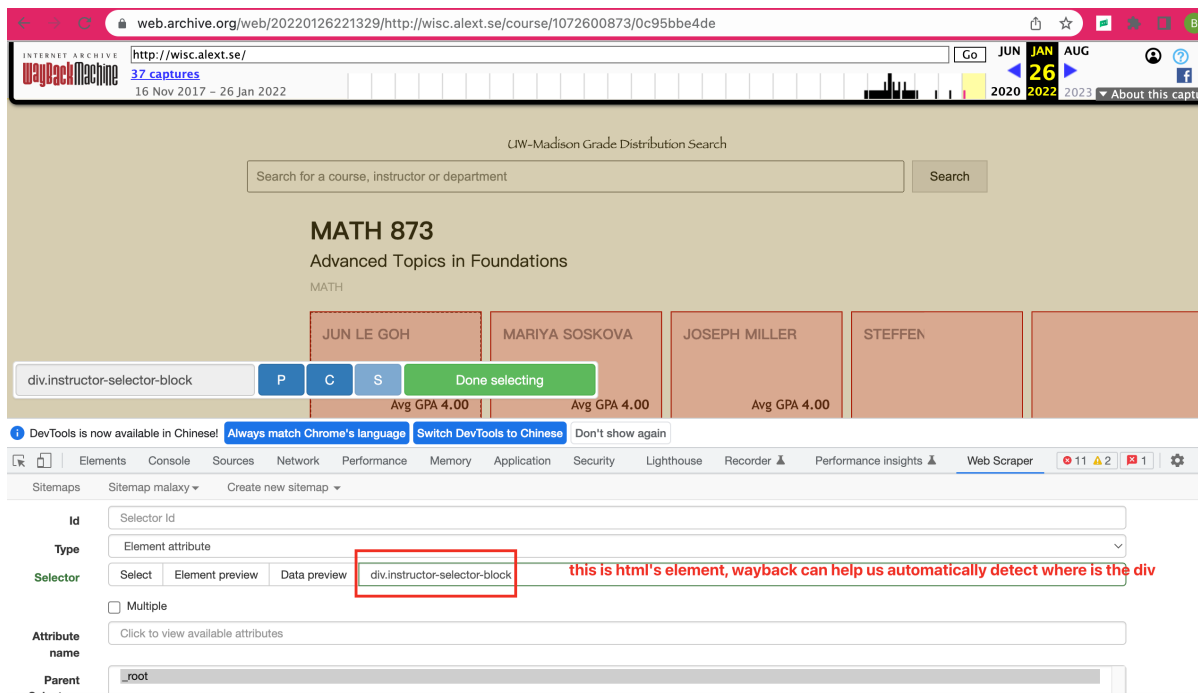

For example:

First create a site map:

Go to the content that we want to scrape:

Select those element we want, we could choose element/ text to scrape:

JUN LE GOH

MATH 873
Advanced Topics in Foundations

MATH 240
Introduction to Discrete Mathematics

COMP SCI 240
Introduction to Discrete Mathematics

div.instructor-course-grid-block P C S Done selecting

Fundamentals of Set Theory

| Elements | Console | Sources | Network | Performance | Memory | Application | Security | Lighthouse | Recorder | Performance insights | Web Scraper |

| Sitemaps | Sitemap malaxygrade ▾ | Create new sitemap ▾ |

Id	courseName
Type	Text
Selector	Select	Element preview	Data preview	

The selector is required and cannot be empty

☐ Multiple

| **Regex** | |
| **Parent Selectors** | _root |

3. In this time I also attached my personal draft code for cleaning

I cleaned the data in this MalaxyDataCleanningDraft.ipynb after every crawl. The raw data is not in a uniform format each time, there is no fixed cleaning procedure. But I share this file and hope it can support any details of my work.