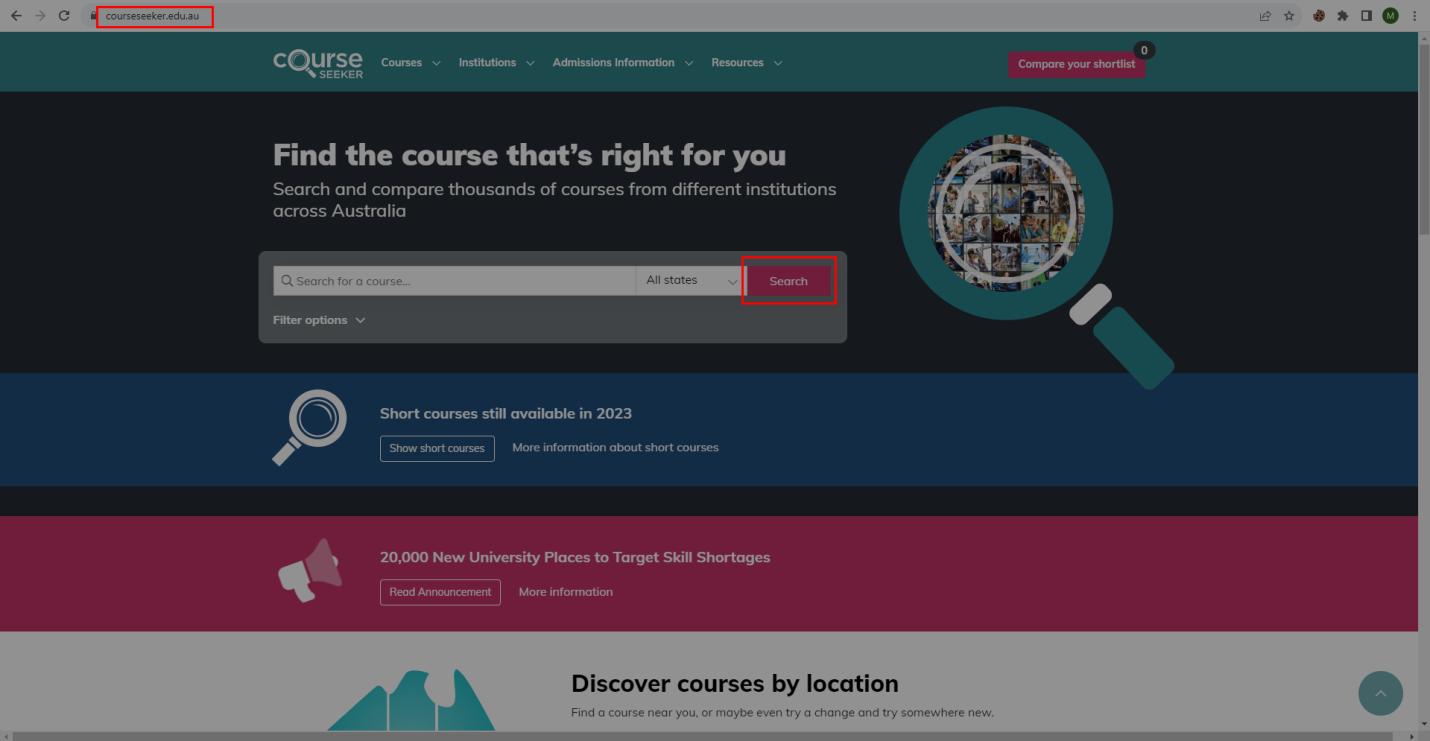
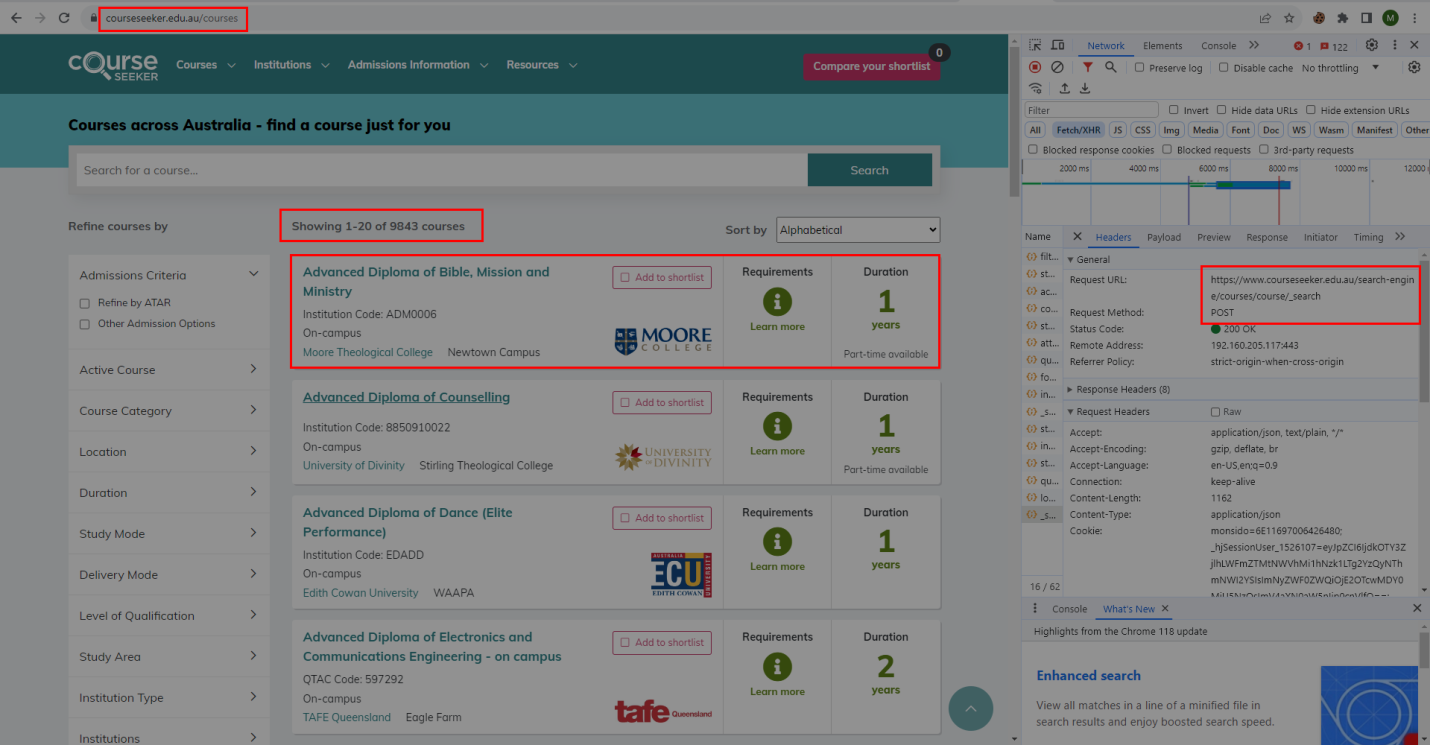
**Course Seeker Documentation**

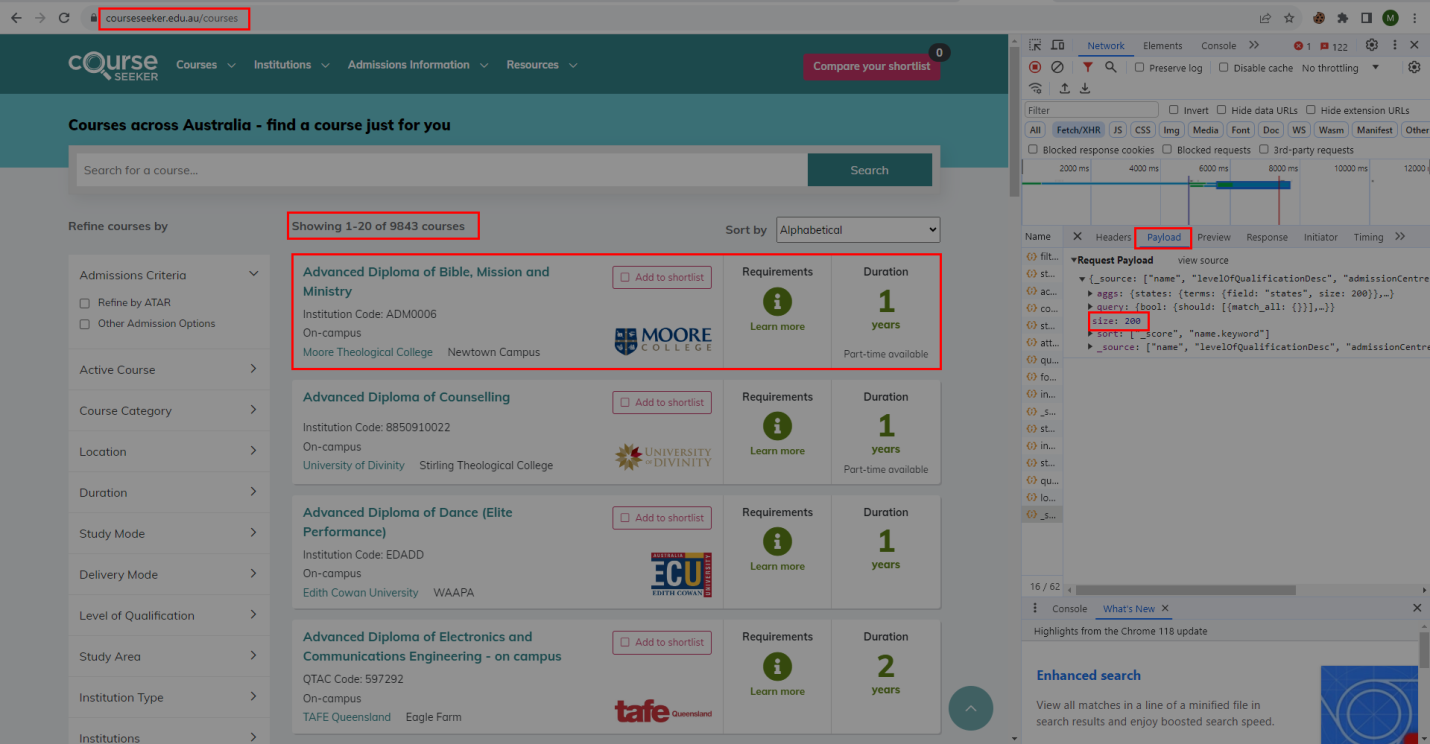
**Table of contents:**

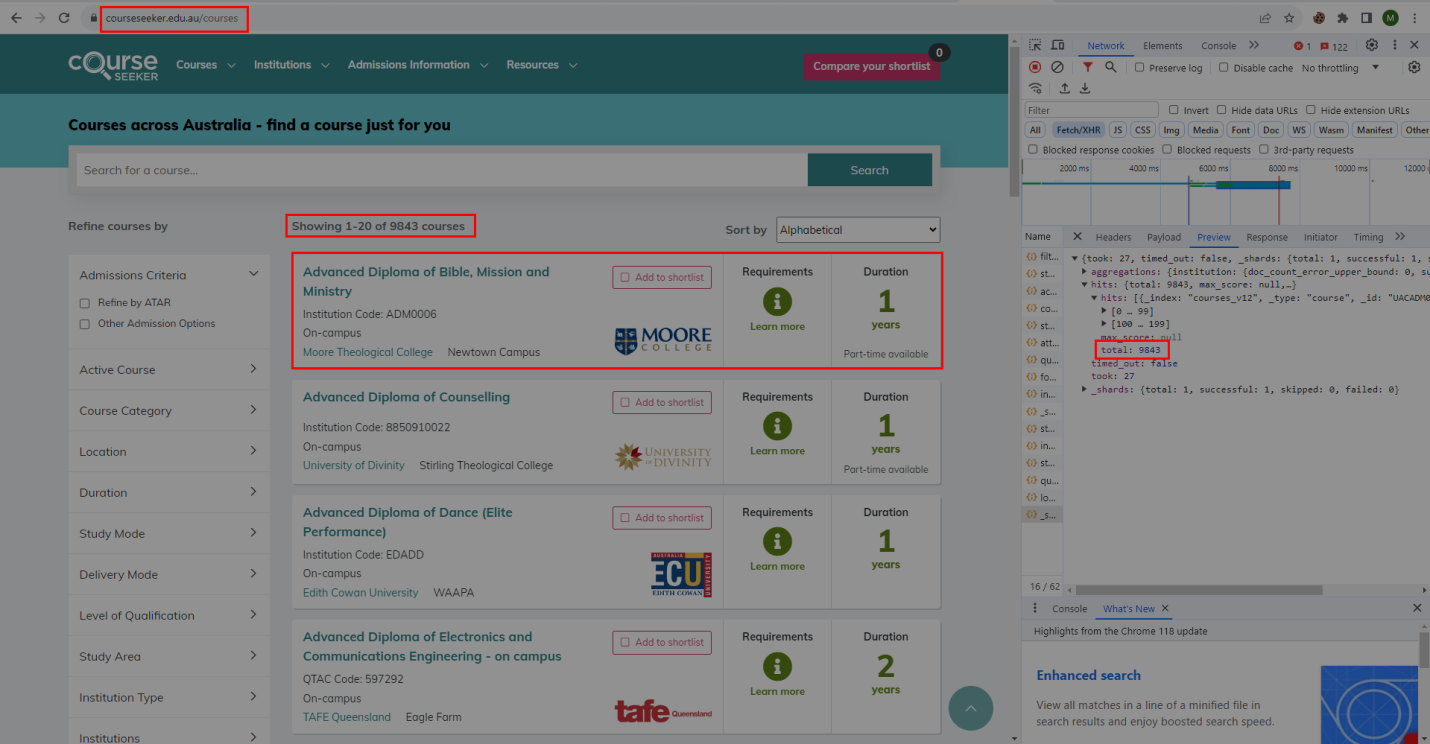
* Get course URL’s
* Save course URL’s
* Get institution URL’s
* Save institution URL’s
* Get course data
* Get institution data
* Merge course and institution data
* Conclusion

**Get course URL’s**

We get institution and course data from course seeker website. Website URL: <https://www.courseseeker.edu.au/> . All course URL’s are presented in API URL. Now we will collect all course URL’s from <https://www.courseseeker.edu.au/search-engine/courses/course/_search> this API URL. This URL method was post method so pass payload while we send request. Change payload size 200 to 9850(total course count). This data are presented in JSON formatted. So we will collect all course API URL ID’s from \_id key and convert to API URL’s. Use **fetch\_course\_api\_url.py** script to get course API URL’s





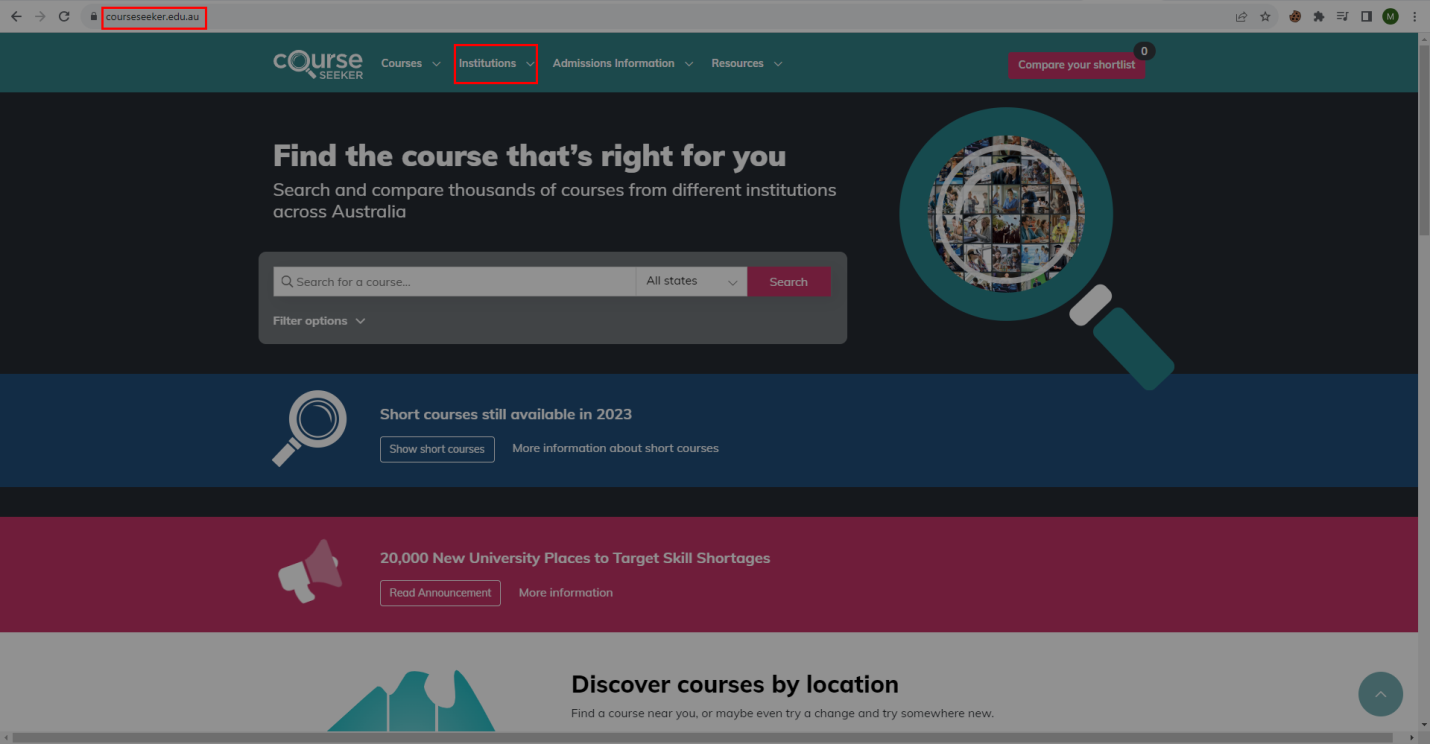


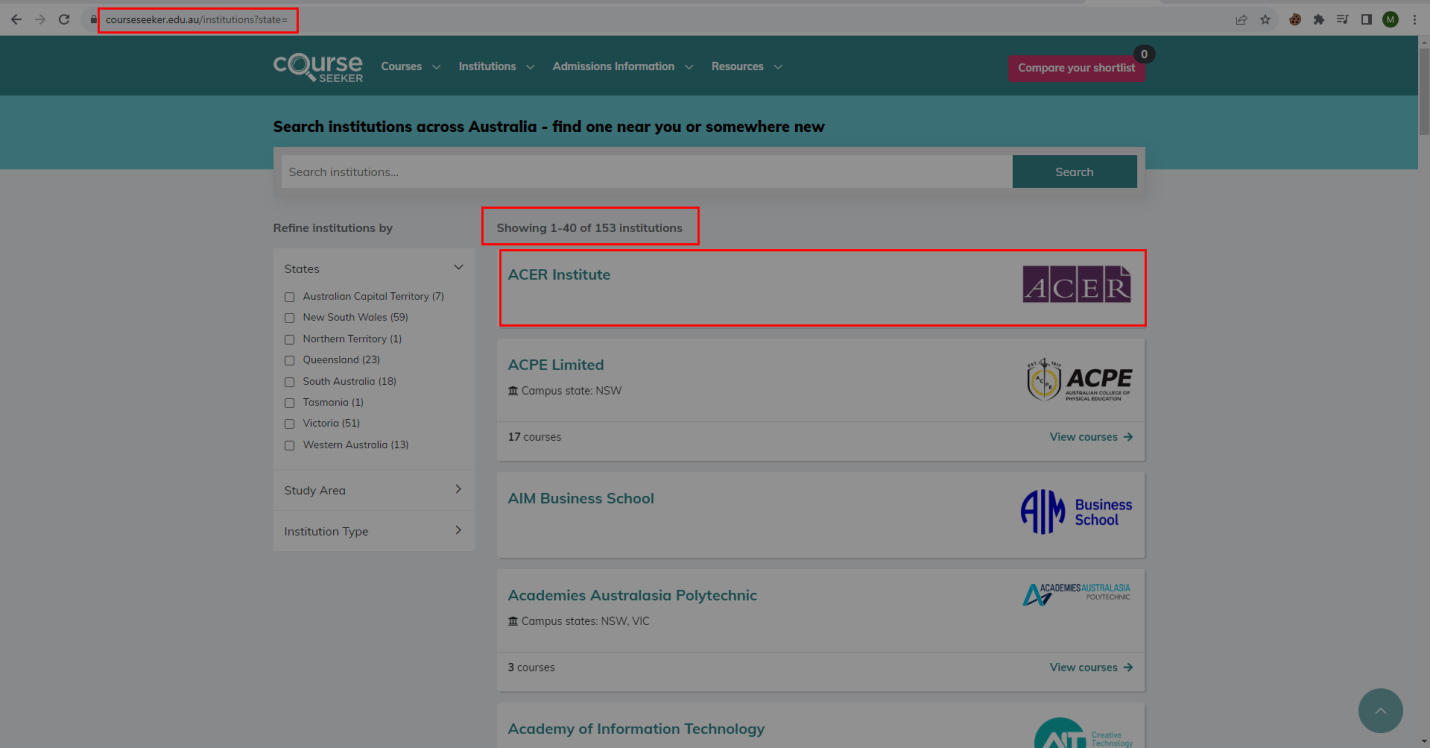
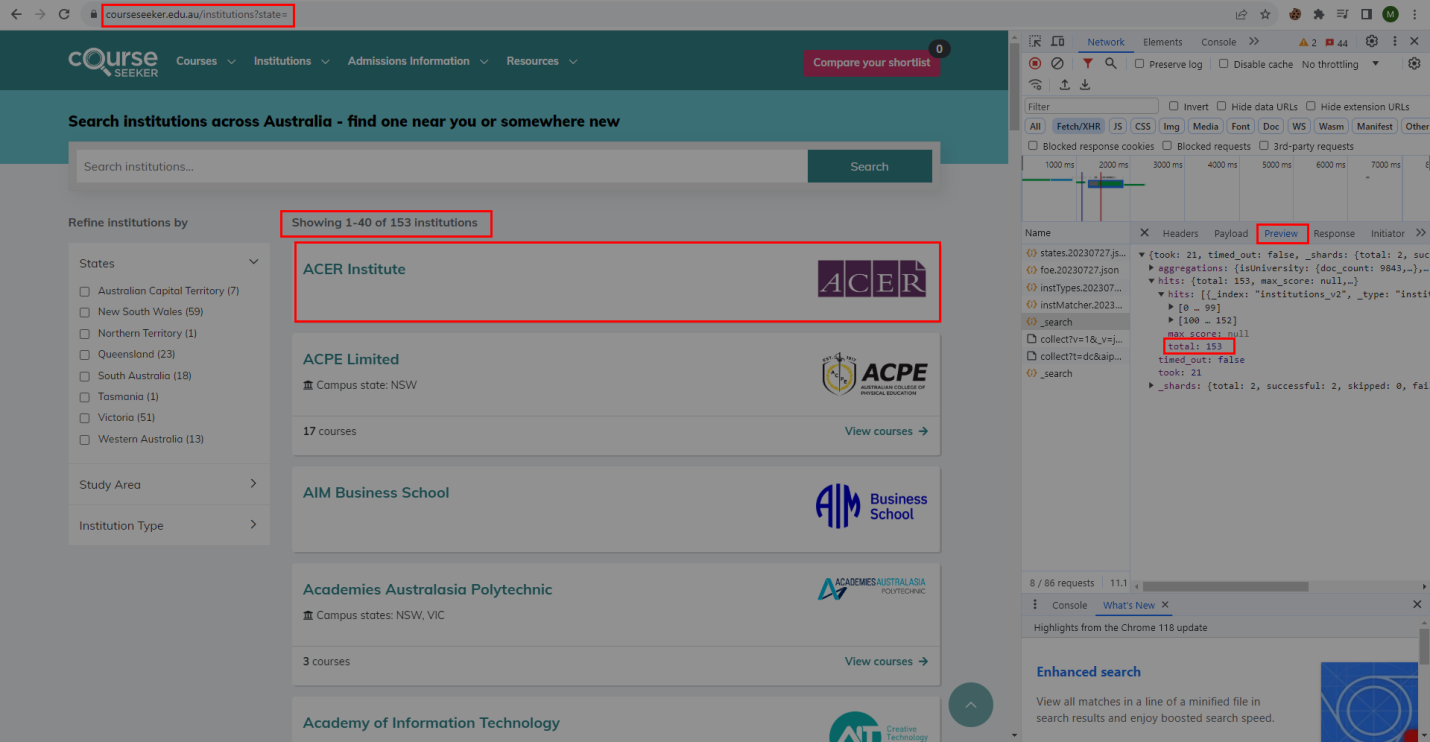
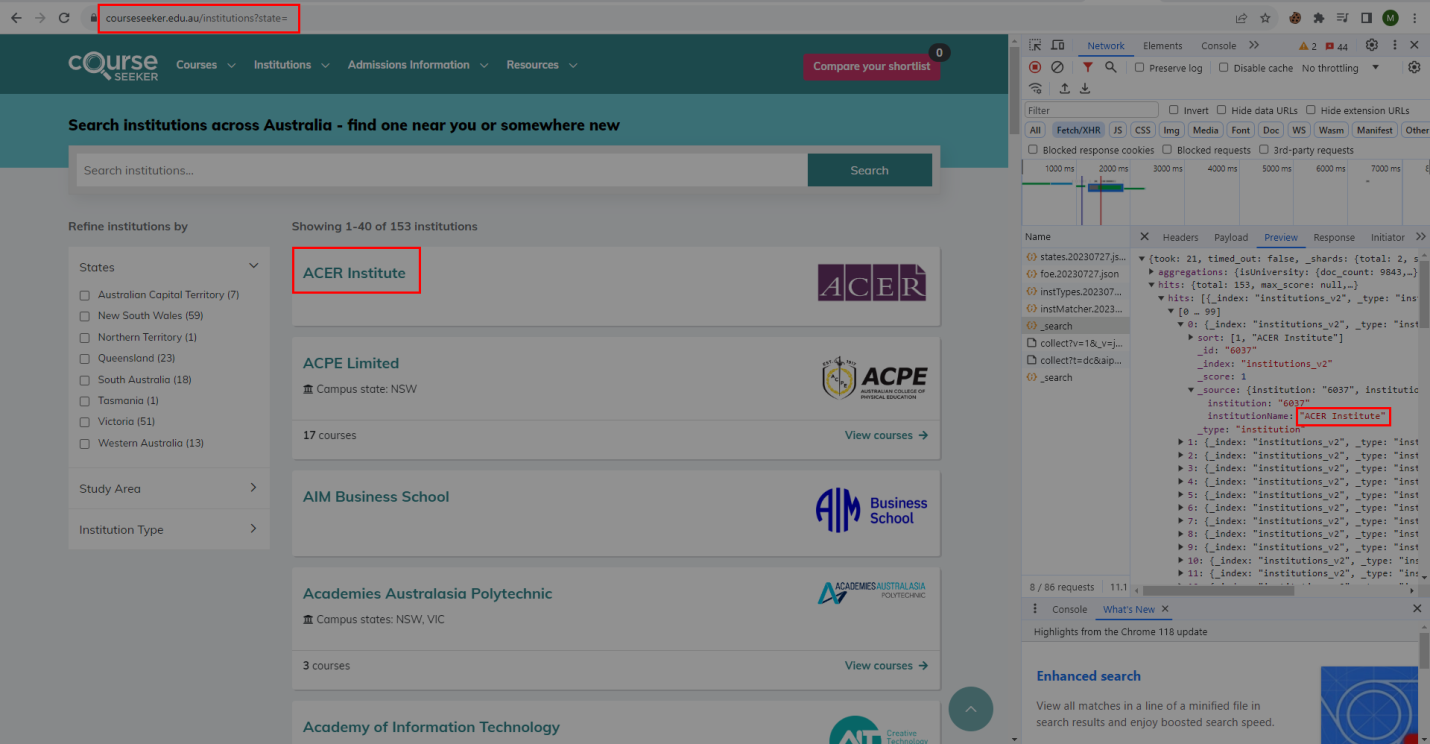
**Save course URL’s**

After got course API URL’s we want to save as a JSON files. Use **save\_course\_api\_url.py** script for save course API URL’s.

**Get institution URL’s**

All institution information are not presented in course files. So we collect all institution URL’s. Go to **course seeker** webpage and move to **institutions tab** and click **all states.** Now we see all institutions details.

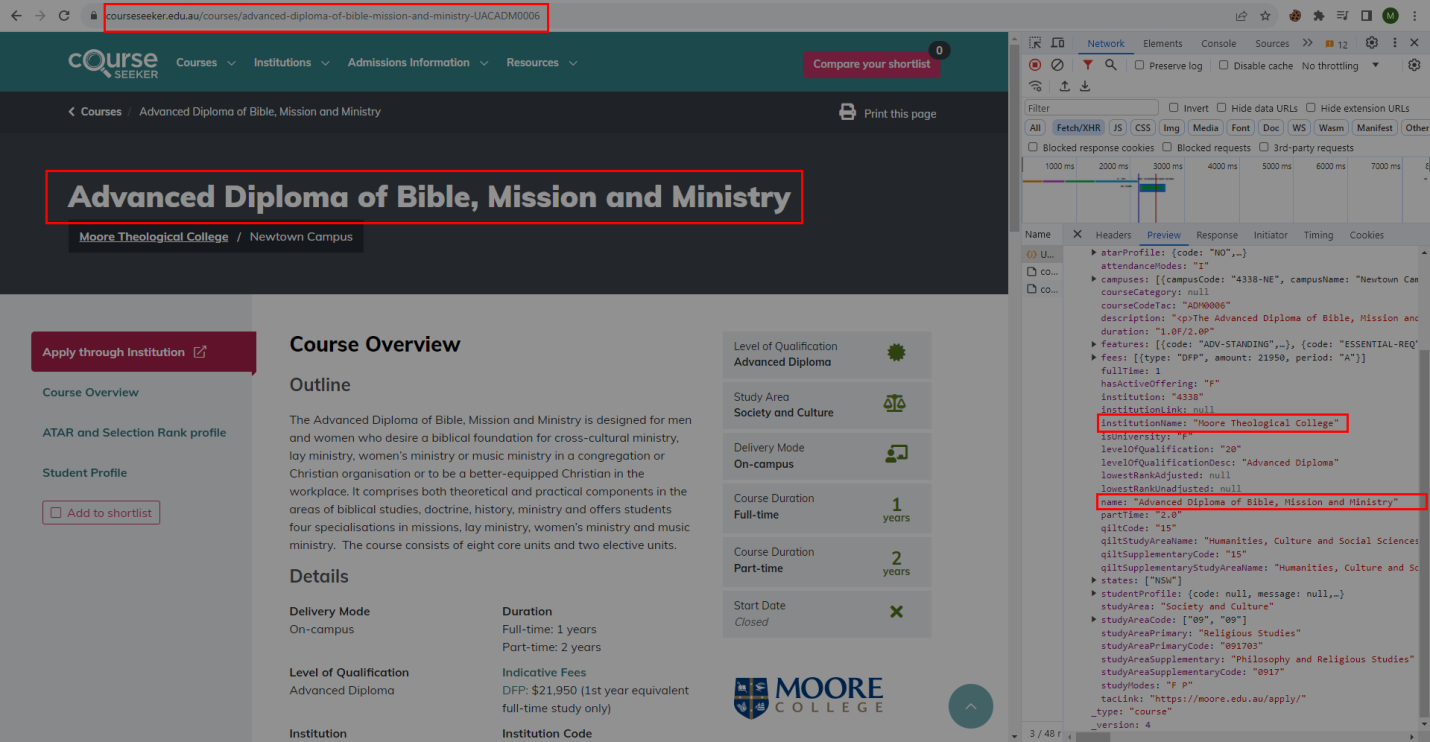


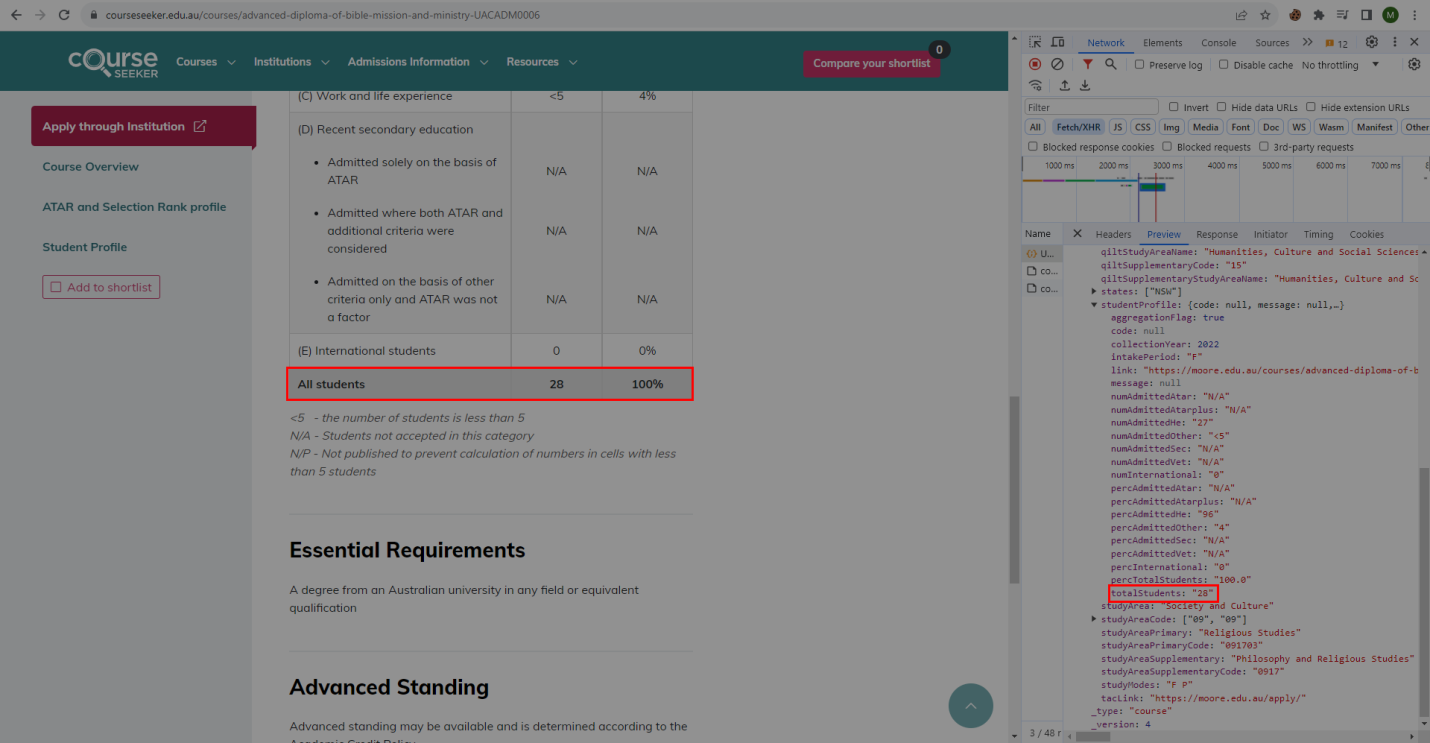
In institution API URL, they are not provided all institutions URL’s. We will convert all institutions name to institutions URL’s. After converted URL’s collect all URL’s. Use **fetch\_institution\_url.py** script to get all institutions URL’s.  

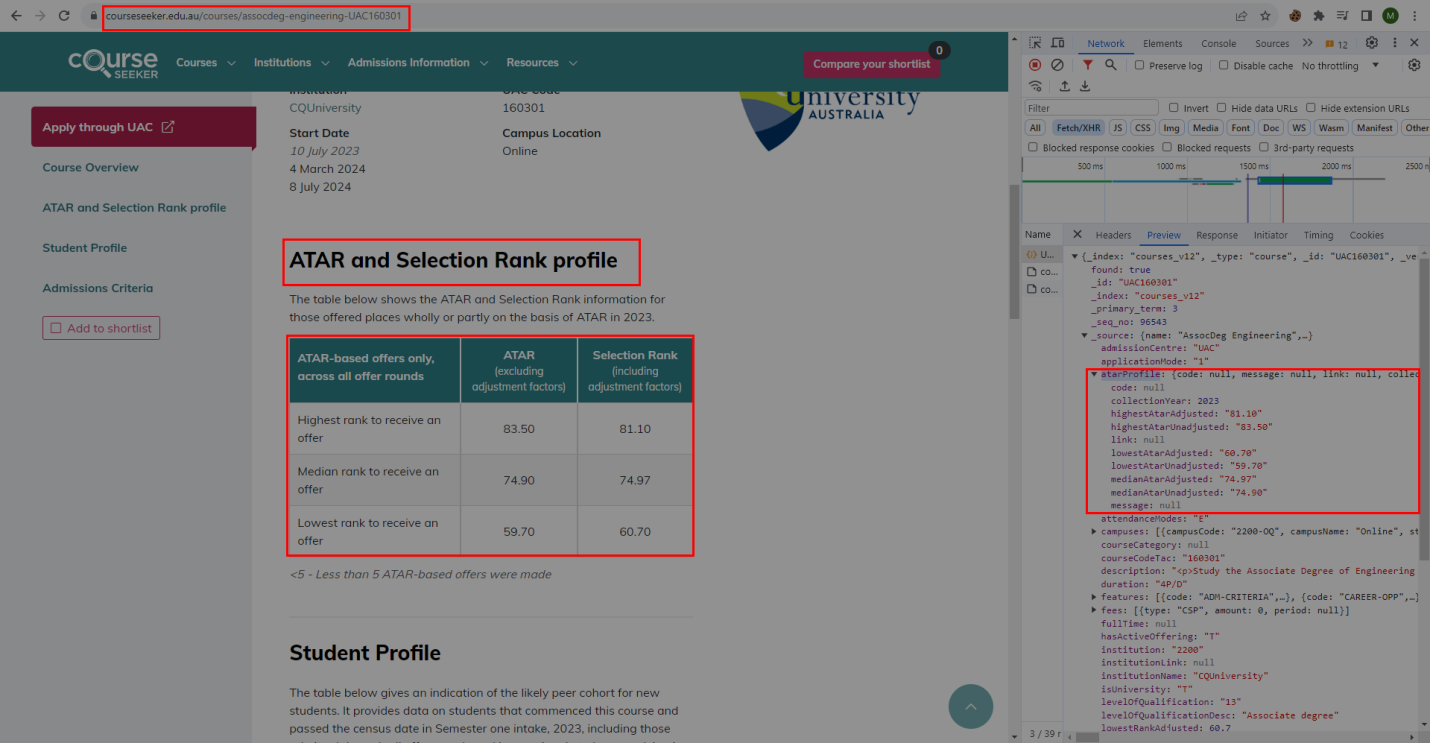
**Save institution URL’s**

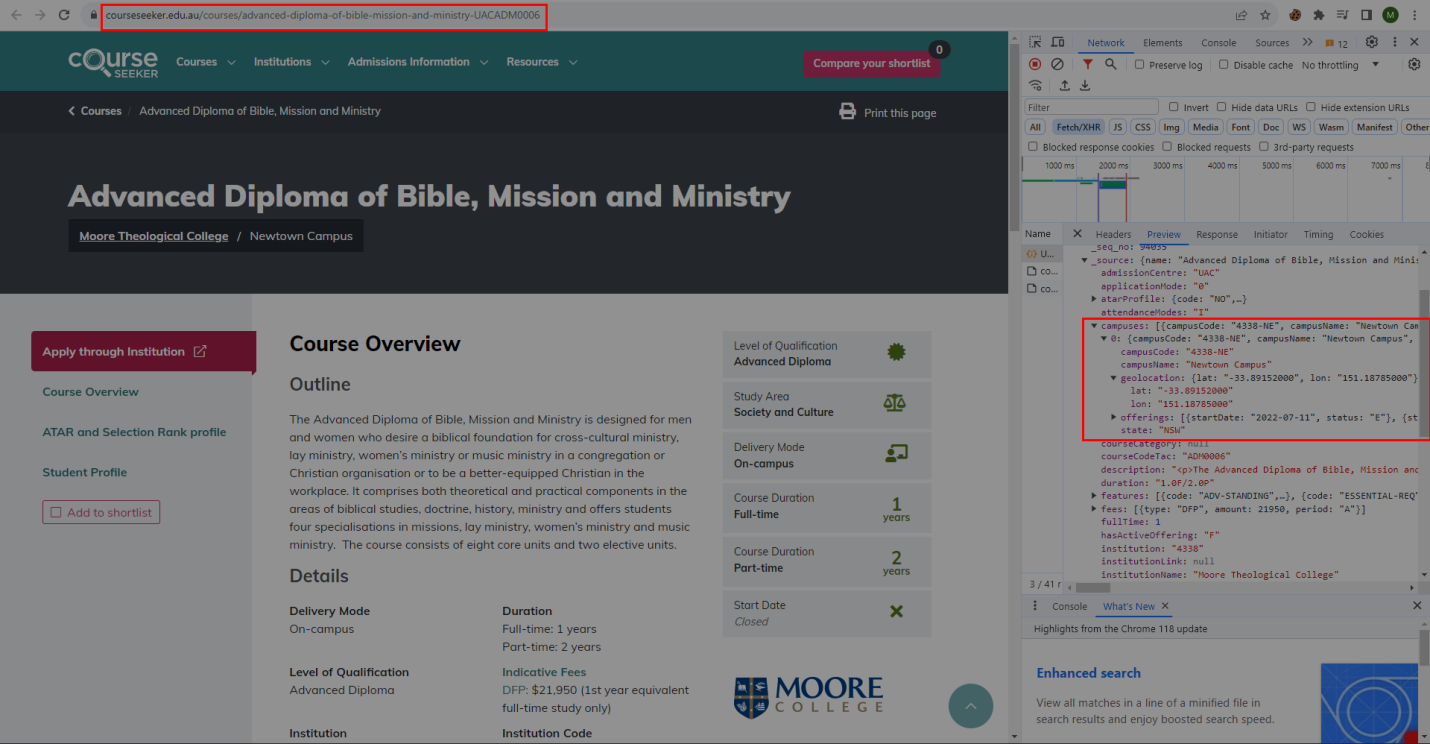
After got institution URL’s we want to save as a HTML files. Institution URL’s are not presented in institution HTML page so insert institution URL and Institution name while saving as a HTML page .Use **save\_institutions\_url.py** script for save institution URL’s.

**Get course data**

After completed course API URL’s saving process get course data from course JSON files. Get course id from **\_id** key. Get institutions URL and course URL from **institutionName** and **name** key**.** After got institutions name and course name convert to institution URL’s and course URL’s. 

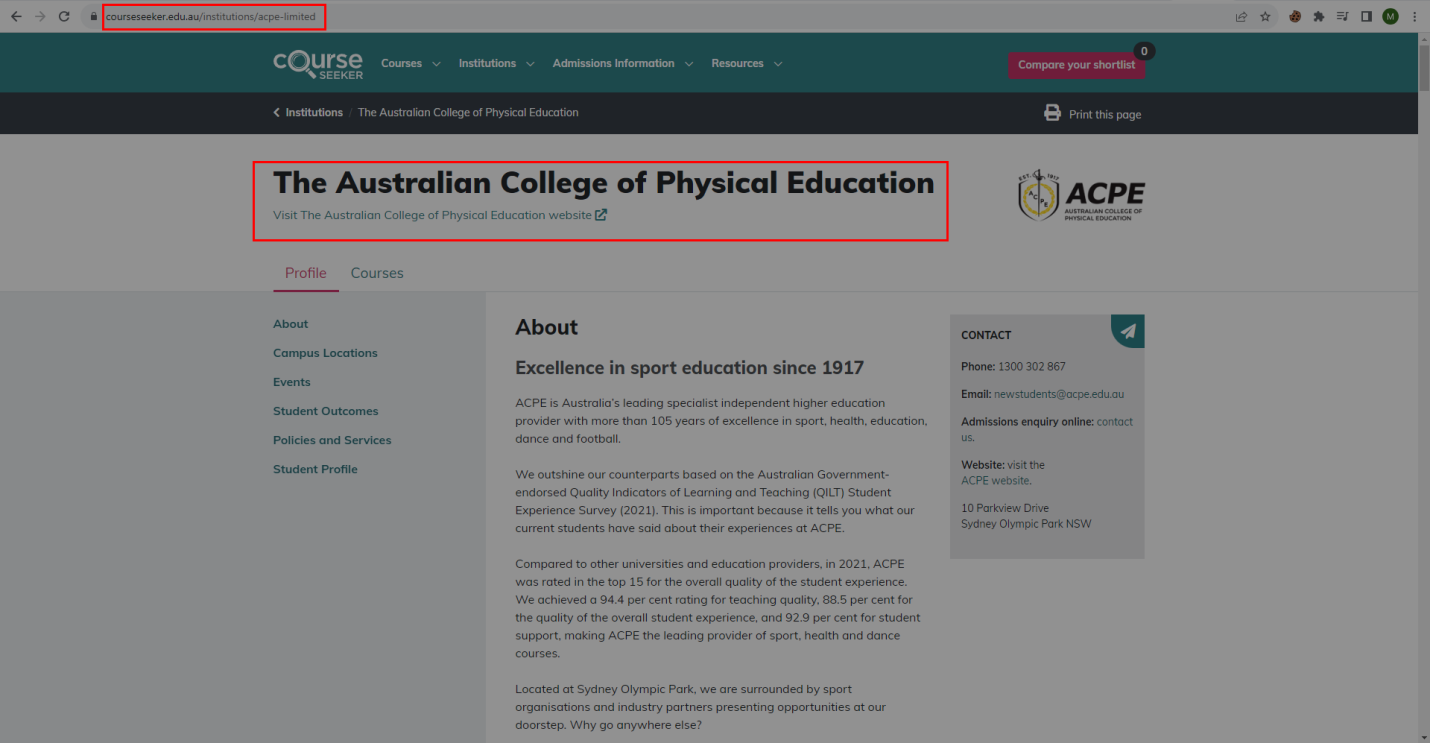
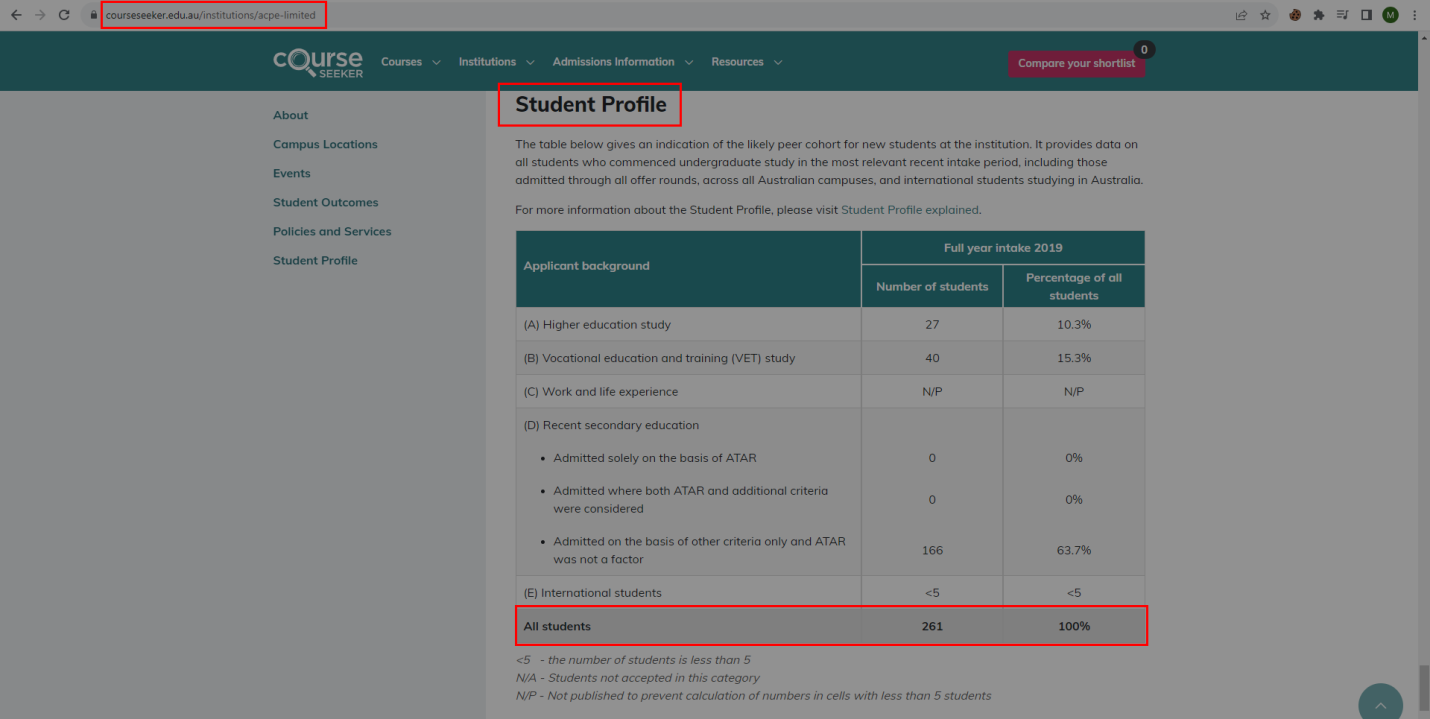
Get total students data from **studentProfile🡪totalStudents** key. 

Get **ATAR and Selection Rank profile** data from **atarProfile** key. 

Get campus data from **campuses** key. 

After got course data create CSV file. Use **fetch\_course\_data.py** script to get course data.

**Get institution data**

After completed institution URL’s saving process get institution data from institution HTML files. Get institution name, institution URL, establishment year and total students count from institutions HTML page. Use **fetch\_institutions\_data.py** script to get institution data. 

**Merge institution data and course data**

Course total students count and campus code are not presented in institution HTML page. So we merge institution data CSV file and course data CSV file. After that create final total students CSV files. Use **merge\_csv.py** script to merge institution data and course data.

**Conclusion**

Now we have all CSV files. This is a whole process for get **Course seeker** data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| URL Type | Total URL’s | Success URL | Failure URL | Permanent Failure URL |
| Course URL’s | 9850 | 9850 | 0 | 0 |
| Institution URL’s | 153 | 153 | 0 | 0 |

|  |  |
| --- | --- |
| Start date | 10/11/2023 |
| End date | 10/16/2023 |
| Total days | 4 |

**🟋🟋🟋**