cmpe 230 Homework 2

Barış Başmak Kayacan Vesek
2016400087 , 2016400114

April 28, 2019

1 Introduction

Goal: Making a Python program called bucourses.py that will crawl Bogazici University's OBIKAS registration pages and extract information about the courses and how many distinct instructors there are offering these courses.

It gets two arguments (year – semester) then returns the courses and some information of that courses between these year-semester pairs including both. (Semester part must start with a capital letter as seen in the project description.)

The courses that were offered in the semesters in the interval have x's in the respecting columns.

In the "Total Offerings" section U and G are the sum of the courses offered throughout the semesters whereas I is the number of distinct instructors.

In the course code section U and G represent the number of distinct courses offered for undergraduate and graduate students respectively.

In the semester columns U and G show the number of undergraduate and graduate courses and I stands for the number of distinct instructors for that semester.

/ # in the total offering column stand for the number of semesters that the course was offered and how many instructors have offered it.

Note 1: Asian Studies and Asian Studies with Thesis have the same code but they are taken as different departments as their names are different.

Note 2: All course names are in quotes because some course names have commas (,) in them which disrupt the csv format.

Note 3: The instructor count of a course does not include instructors who only appear in the ps and labs.

Note 4: It has a shebang line to open program directly by writing ./bucourse.py (without python3 at the start) If you run the script from the shell, it will inspect the first line to figure out what program should be started to interpret the script. "#!/usr/bin/python3" our interpreter is located here so if you want to use program as exacutable you have to have interpreter located that specified location.

2 Methods

- **getTime:** Tokenizes the arguments for the URL links
- **getLessonCode** Removes the section code from the lesson code and returns the sectionless code and if it's an undergrad course the method returns 1, if it is an graduate course it returns 0 as a second tuple
- **linkertuple** Merges the course codes and course names in a list of tuples while keeping the list order.
- **getPage** Gets the page for the specified year, semester, department code and department name. As a parameter it has a variable mp which is dictionary of sets. It keeps the instructors names' set for every distinct course code (mp[course_code]=set(instructors) format)
  First, it takes the page data by using request.get method from requests library, after then by using BeautifulSoup it process data.
  After than it returns course_list lesson_list instructor_set for that specific course code , year, and semester given as parameters.
- **UorL** Returns 0 if the course is an undergraduate course, 1 if it's a graduate course.
- **solveDep** It gets the interval and the department then prints to stdout, the wanted values in the csv format. This method does the all operations, first it calls getPage the take the processed data. Then it finds the cource counts, total offerings, and cources etc. In the end it prints the all values.
- **main** It calls solveDep function for all the departments with Dept. code in order of alphabetical order.

3 Summary

Our code gives the desired result successfully in the specified time limits given that the internet connection is sufficient.