# Homework 2: XML & XPath

## DSCI 551 – Fall 2023

## Due: 11:59pm, September 29, 2023

## Points: 100

Similar to homework 1, in this homework assignment, you are asked to create an app for managing courses. However, you are required to store the data of the app in an **XML** file called hw2.xml (case sensitive). In other words, you can regard hw2.xml as the database of your app. Similar to homework 1, the database (the XML file) should store the following information about students, instructors, courses, and their relationships.

* Student: id (e.g., “s100”), name (e.g., “John Smith”), and program (e.g., “Applied Data Science”).
* Instructor: id (e.g., “i100”), name (e.g., “Bill Gates”), and department (e.g., "Computer Science”).
* Course: number (e.g., “DSCI 551”), title (e.g., “Foundations of Data Management”), and semester (e.g., “Fall 2023”).
* Relationships: students take courses, and instructors teach courses.

You are free to design the structure of hw2.xml for storing the above content.

Please complete the following tasks. All the tasks assume that your database is stored in hw2.xml.

1. [50 points] Write the following Python programs for adding data into the database. Note that after the execution of each program, your hw2.xml should have updated data.
   1. add\_student.py:

Execution format: python3 add\_student.py “s100” “John Smith” “Applied Data Science”

* 1. add\_instructor.py:

Execution format: python3 add\_instructor.py “i100” “Bill Gate” “Computer Science”

* 1. add\_course.py:

Execution format: python3 add\_course.py “DSCI 551” “Foundations of Data Management” “Fall 2023”

* 1. take\_course.py: which student takes which course in which semester. Note that your program needs to check the validity of input. That is, student id, course number, and semester indeed exist in the database. Report an error otherwise.

Execution format: python3 take\_course.py “s100” “DSCI 551” “Fall 2023”.

* 1. teach\_course.py, which instructor teaches which course in which semester. Similarly, you need to check the validity of input based on the data in your database; and report error otherwise.

Execution format: python3 teach\_course.py “i100” “DSCI 551” “Fall 2023”

1. [50 points] Write Python programs for the following queries.
   1. [25 points] find\_student\_courses.py: it takes a student id and returns student name and a list of courses taken by the student. For each course, report its course number and semester when the course was offered. Return the output in XML format.

Execution format: python3 find\_student\_courses.py “s100”.

* 1. [25 points] find\_instructor\_courses.py: it takes an instructor id and returns **instructor** name and a list of courses **taught** by the instructor. For each course, report its course number and semester when the course was offered. Return the output in XML format.

Execution format: python3 find\_instructor\_courses.py “i100”.

Permitted libraries: lxml, sys.

**Submission Instructions:**

1. Please submit 7 .py files and a hw2.xml file (base xml template which should only include initial parent nodes without data).
2. Please do not submit any other files except the 8 files.
3. Your code should not create any extra file. All the modifications (add/delete) should be done in the hw2.xml (Your database for this project).
4. If your code does not run, you will get 0 points.