Homework 2: XML & Distributed Storage
DSCI 551 – Spring 2024
Due: 11:59pm, February 16, 2024, Friday
Points: 100

In this homework, we will use XML files for distributed data storage. Suppose we are to store the information about a collection of books. Each book is represented using an XML element, containing author, title, year, and price. Suppose we partition the books by hashing on the author of the book and store books in two different XML files, depending on the hash value. For example, if h('John Smith') = 0, we will store it in file0.xml; otherwise, in file1.xml.

The homework consists of the following tasks. You need to create two XML files of your own and design an appropriate hash function for hashing on the author names.

Complete the homework by writing the below methods in the given template.

Operations:

1. [40 points] **add_book**: Takes a JSON object that contains the book information (with values for all four attributes) and stores the book in one of the XML files, based on the hash value of the author as described above.
   Example:
   > python3 script.py add_book 100 '{"author": "john smith", …}'

2. [30 points] **search_by_author**: Finds all books written by the author.
   Example:
   > python3 script.py search_by_author "John Smith"

3. [30 points] **search_by_year**: Finds all books published in the specified year.
   Example:
   > python3 script.py search_by_year 2023

ALLOWED LIBRARIES: sys, json, lxml

Some resources for lxml:
1. https://lxml.de/tutorial.html
2. https://lxml.de/xpathxslt.html

SUBMISSION INSTRUCTIONS:
1. A single python file with name: [Student_Name]_hw2.py
   - replace Student_Name with your name Eg. **John_Smith_hw2.py**
2. Do not modify any contents in the template. Just fill the template by reading the comments.
3. The test script will accept the return data as specified in the template.
4. Testing is done by a test script with different test cases. So points will only be awarded if the method returns the expected result.
5. You will get 0 points if the code breaks for any syntax errors or any other problems. Please test the code thoroughly before submitting.