1. Using the XML document books.xml provided with this homework, define the following queries in XQuery (20 pts):

   (a) Give the titles of all Books sorted by price.

   (b) How many books were written by Rajati?

   (c) Give for each author, the number of books they have written.

2. Consider the data set FoodServiceData.csv about the inspections of food services in some cities (mainly Louisville) in Kentucky. The format of the data set is csv (comma-separated values)(50 pts).

   (a) Provide a Python script "xmlizer.py" that coverts the data set into an XML file using pretty print from lxml library. The root element should be <foodservices>and the tag for each food should be <foodservice >. For example:

   python xmlizer.py FoodServiceData.csv FoodServiceData.xml

   where FoodServiceData.csv is the provided data set and FoodServiceData.xml is the output XML file name. Note that your script will be tested using similar data sets (with the same format).

   (b) Write a Python script "enum.py" that takes the XML file you produced in 2a and yields the number of foods ervices by their Type Descriptions. The output should be in ascending alphabetical order in terms of location types, and each type should be in a separate line. Use whitespace to separate type and count.

   Example: python enum.py FoodServicesData.xml

   Example output: CATERERS 125

   FOOD SERVICE 500

   (c) Repeat 2b for grade of the food services, i.e. write a script that shows how many food service of each grade (A, B, C, etc.) the data set has. What happens to the food services with missing grades?

3. In what follows, you will learn how to scrap webpages using XPath. In particular, consider search results from the website Thriftbooks.com. You are provided with HTML files of search results on books about XML, Big Data, and Machine Learning. (50 pts)

   Write a Python script "dig.py" that takes a search result html file (e.g., "book.html" below) and extracts the title, price, format, condition, and authors of the books. The script should output an XML file containing the extracted information with root element <books>. Each book should be represented as an element book and the title, price, format, condition, and authors of each book should have their own tags. Use the lxml library.

   Example:

   python dig.py book.html book.xml

   Hint: Set encoding option as "utf8" in open() if you encounter UnicodeDecodeError when opening html files. Use etree.HTML() instead of fromstring() for html files to avoid lxml.etree.XMLSyntaxError.

Also, make sure that non-book items are not the result.