CSCI 3020 ASSIGNMENT 3 (50 POINTS)

DUE TUESDAY, OCTOBER 6, 11:59 PM

GOALS

This assignment has you extract data from an XML document using a SAX parser.

GETTING HELP

- Post general questions on the discussion board in D2L. Please refrain from posting more than 1 line of code, otherwise your post may be removed.
- Send me email, nicholsonja@apsu.edu, or come to my office. I'll be happy to help. If emailing me, please do not send me a code snippet; send your entire .cpp so that I can compile and test it.

BACKGROUND

The pesky librarians at the Small Town Library is at it again. This time they want some data simplified. They have a file, **mondial-3.0.xml**, and they want three files created from this data. The three files will be used by the librarians to help library patrons.

DATA FILE FORMAT

The mondial-3.0.xml is included with the ZIP for the assignment. The file is contains a subset of the data from the Mondial Database, which is created and managed by the Institute for Informatics at Georg-August-Universität Göttingen (http://www.dbis.informatik.uni-goettingen.de/Mondial/).

It is a valid XML in a UTF-8 encoding that is compatible with ASCII. You should study the elements in the file so that you understand its basic structure.

MAIN REQUIREMENTS

Your program should read the mondial-3.0.xml file and create three CSV files.

1. **country.csv**

   This file holds the names of the countries.

   Although it is a CSV file, it will only have one column – the name of every country in XML. Each country name should appear on a line by itself. For example, the first lines in the CSV file will be:

   Albania

   Andorra

   Austria

2. **cities.csv**

This file should list all the cities

Each line of this file will be a city name followed by the country name. They should be separated by a comma. Each line represents one city/country pair. For example, the file starts with Albania. The first few lines in the CSV file will be:

Tirane, Albania

Shkoder, Albania

Durres, Albania

3. **religions.csv**

This file should list the primary religions in each country.

Each line should start with the country name and then a list of the religions for that country each separated by commas. For example, the first three lines would be:

Albania, Muslim, Roman Catholic, Albanian Orthodox

Andorra, Roman Catholic

Austria, Roman Catholic, Protestant

You do not have to sort any of the data. Output it in the order you find it in the XML file.

You do not have to prompt the user for anything. Read in the mondial-3.0.xml file and create the three required files.

You must use the Xerces library, which is demonstrated in lecture and the videos.

TIPS

- Use the code from the videos on Xerces as the starting point. Don't write this from scratch.

ADDITIONAL C++ REQUIREMENTS

Additional requirements

1. You may not use any global variables.
2. You may not use cstring, that is, do not do this:
   #include <cstring>
   You can use string, that is, this is ok:
   #include <string>
3. You may not use the **exit()** function.
4. Except for items #1, #2, and #3 above, you may use any feature of C++ defined in the language or STL. Your CSCI 1010/2010 book covers basic C++.
5. You may not use any non-standard C++ library or feature specific to your compiler. For example, Visual Studio users may not use the XML libraries provided by Microsoft.

6.   You must include a comment at the top of your main.cpp file in this format:

```
/*******************
John Nicholson
CSCI 3020 Section 09
Fall 2015
Assignment 1

Programmed on Windows 7 using Visual C++ 2010 Express

This program converts ...
*******************/
```

If this comment is missing, you will receive 0 points on the assignment. You should obviously change the name to your name, the section to your section, the system to the system you used, and write a real description of the program. You must include the specific operating system and the compiler/IDE you used.

Please note that **Windows** is not an operating system; **Windows 7**, **Windows 8**, and **Windows 8.1** are operating systems. **Visual Studio** is not a compiler/IDE. **Visual C++ 2010 Express** and **Visual Studio Ultimate 2013** are compiler/IDEs. You should be specific about which OS and compiler/IDE you used, otherwise you may lose points in grading because I graded using the wrong compiler.

NOTE: I will not regrade if you do not put in the correct comment or are not specific enough, and I end up testing your code with a different compiler than you did causing your code to crash.

SUBMITTING

Upload the following three files to the Assignment 3 DropBox in D2L:

- Your main.cpp file. Do not upload the xerces library. Just your main.cpp
- The three CSV files generated by your program.

Do not upload your entire Visual Studio project (or project of whatever you are using). Just upload the three files listed above.