Week 8:Reading Webpages

# Lab 14

## IMPORTANT INFORMATION:

You have been provided with two files included in the zipped C# version. Use these programs to help you complete your project.

(Note: Some of the requirements specified below have been completed for you)

## Concept Summary:

1. File I/O
2. Multi-threading

# Submission Guidelines:

You will turn your program code (Java or C#) and your response in a Word or pdf format.

# Description

In this lab you will write two programs. One using standard programming constructs and the other using multi-threading. Using print statements (as shown in the tutorials) justify the idea - multi-threading improves processing time.

When you downloaded a page from the internet you accidently include the HTML tags.

Such as…..

In order to save time removing these tags manually you decide to write program that might be useful again in the future which will remove the tags and leave just the text.

You are required to do the following:

1. Create a class that takes care of reading the file content into a list and splitting the list into multiple lists in order to prepare the workload for the threads. This class should include two methods:

* A method *ScanList()* should accept a Scanner object and read the entire contents of the file into the list
* A method *SplitList()* should split the contents of the list into multiple lists depending on the number of threads you intent to use.

1. A class that contains a processing method which strips the content between <> from a certain list
2. A driver program that attempts to perform the stripping operation with and without threads. The driver program should make use of some sort of timer in order to track the time required for executing with or without threads. Do you notice the difference between these two? Why do you think this is happening?
3. Please note that the driver program does not modify the original file but rather employs data structures in order to compute the time required to strip the content between <> from the file.
4. You can assume that a tab is any text between the characters < and >. You can also assume that the file is a well-formatted HTML document and that there are no < or> characters inside tags.

**Sample Code:**

### Thought Question:

**Write a short response (not more than 3 sentences or 100 words) explaining the improvement in time taken for processing data when using multi-threading.**.

# Grading

100 %: Attempted lab and submitted fully functioning lab exercises (all requirements met), with complete headers and clear I/O statements before due date. Also, thought question was answered correctly. Example header should look somewhat like the following:

//Name:

//Class name:

etc.

95 %: Attempted lab and submitted fully functioning lab exercise (all requirements met) but incomplete headers and/or unclear I/O statements before due date. Also, thought question was answered correctly.

90%: Program compiles and stripping computation with both threads and without threads have been successfully implemented but major logic or syntax errors are present. Also, thought question was answered correctly.

80%: Program compiles and stripping computation with either threads or without threads has been successfully implemented. Also, thought question was answered correctly.

70%: Program compiles and stripping computation with either threads or without threads has been attempted.

0 %: Did not attempt lab or did not submit it before the due date.