**CSCI 232 Operating Systems**

**Homework 1 – Shell scripts**

This assignment will introduce you to a shell scripting language, **bash**, the most commonly used shell at the moment. The language is useful for manipulating files, performing search and data collection. Your task is to create a script that will analyze several text files in order to perform primitive automated plagiarism checking. This assignment, as always, is individual work.

Imagine you are a teaching assistant who wants to compare the solutions of everyone in your class to some solution you found on the Internet.

* Assume the student solutions are a set of C files stored in a **tar** archive that has been compressed with **gzip** (very common on unix systems), and the solution to compare them with is a single C file.
* Assume also that you want to do a quick initial check for students who gave credit to their source, as required.

The task is to write a script that will accept three command line arguments specifying:

1. the filename of the archive containing the student submissions (e.g., submissions.tar.gz)
2. the filename of the solution to compare them with (e.g., solution.c), and
3. the keyword to search for in case the student has cited the source (e.g., stackoverflow)

An example call to the script would be:

**./hw1.sh submissions.tar.gz solution.c stackoverflow**

The script should perform the following actions:

1. Unzip and extract the contents of the archive
2. Create three folders named, exactly:
   1. “Exact copies”
   2. “Referenced”
   3. “Original”
3. Examine each of the student files and place them in the appropriate directory:
   1. Files with no difference from the solution and that do not contain the keyword go in the “Exact copies” folder.
   2. Files that have a reference as specified by the keyword argument should be put into folder “Referenced”.
   3. Files having neither of the above conditions should be put into the folder “Original”.
4. Create an output file named report.txt. It should have four lines: the number of exact copies, the number of files containing references, the number of files containing original work, and the ratio of original submissions to the total number of files. For example:

Exact copies 4

Referenced 4

Original 2

Ratio 0.2

1. Archive and compress the 3 folders, the comparison file (the script’s second argument), and report.txt into a file called results-*current\_date*.tar.gz , where current\_date is the actual date the archive was created.
2. Assure that none of the files/directories used in the processing of the script remain when it exits, only the generated archive should remain. Even the original archive (the script’s first argument) should be removed. Other files that happen to be in the current working directory should not be disturbed.

Other remarks:

* For this assignment, the input is guaranteed to be valid. No error checking is required for things like: correct number of arguments, correct file types, correct permissions, etc.
* Scripts that cannot be run on Ubuntu will automatically receive lowest possible grade. You will have to meet the TA to demonstrate your script in live grading, and a penalty may be assessed if the problem is not trivial.
* If you have made any special assumptions or have specific notes/instructions about your script, please include a short README file with your script containing the explanations.