Assignment Prefix: Lab10

Points: 100

Due Date: Friday, March 31 @ 11:59pm

This is an individual assignment.

**Problem:**

Consider a collection of files that represent stock quotes from stock exchanges. These data files span a number of years and have the two distinct file formats. Prior to 01/01/2012 the file format was:

<ticker>,<per>,<date>,<open>,<high>,<low>,<close>,<vol>

AA.P,D,20110103,75.75,75.75,75,75,1400

AADR,D,20110103,29.88,30.06,29.88,30.06,1500

AAU,D,20110103,4.85,5.24,4.66,5.08,1184300

Starting on 01/01/2012 the format became:

<name>,<ticker>,<date>,<open>,<high>,<low>,<close>,<vol>

Alcoa Inc Pf 3.75,AA.P,20120102,81.5,81.5,81.5,81.5,0

Wcm Bny Focused Growth Adr ETF,AADR,20120102,28.11,28.11,28.11,28.11,0

Almaden Minerals,AAU,20120102,2.52,2.52,2.52,2.52,0

In order to make data analysis easier it is desirable to convert the old file formats to the new file formats. In order to do this it will be necessary to look up name/description associated with each ticker/symbol and then rewrite the data file in the new format. A list of ticker/symbols and the associated name/descriptions are available in separate files that have the following format:

Symbol Description

AA.P Alcoa Inc.

AADR WCM BNY Mellon Focused Growth A

AAU Almaden Minerals Ltd.

On the surface, this appears to be a simple task. For each line in the quote file, find the corresponding row in the symbol-description file and the write out a new file with the desired data fields. What makes this task challenging is the amount of information that needs to be processed. For 2011 there are 1,560 quote files that contain a total 4,247,013 quotes, and there are six symbol-description files that contains 21,882 entries.

**Task:**

To make the name lookup as efficient as possible it would make sense to store the name/descriptions in a hash table.

Your task is fine tune the cyclic-shift hash code computation (presented on page 414 of the textbook) to determine the best shift amount for the ticker/symbol key value. Your program should generate output similar to Table 10.1 on page 414 of the textbook.

Make sure that your results are in a “nicely” formatted ASCII table. See example below.

You do not need to convert any data records or write any output files. You just need generate a table that will allow you to determine the best shift amount.

The necessary symbol-description files are in the file named EODdata-Symbols-20110305.zip.

**Turning in your assignment:**

* **Make sure that all of your code is properly documented.**
* Turn in your assignment using the standard method.
* Copy and paste each of your Java files into the document.
* Paste the screenshots showing the complete output of a complete run of your program after the Java code in your document.
* Export your NetBeans project to a zip archive.
* Turn in the Word document and zipped project as to separate files in a single Blackboard submission.
* You do not need to turn in your data files. We will test your program with a standard set of test files.

+---------+-------------------+

| | Collisions |

+---------+-------------------+

| Shift | Total | Max |

+---------+-----------+-------+

| 0 | 20,375 | 523 |

+---------+-----------+-------+

| 1 | 16,508 | 92 |

+---------+-----------+-------+

| 2 | 10,123 | 123 |

+---------+-----------+-------+

…

+---------+-----------+-------+

| 14 | 12,321 | 23 |

+---------+-----------+-------+

| 15 | 15,321 | 308 |

+---------+-------------------+

This example is for formatting purposes only.

The numeric values will probably not match your results.