

Assignment 1

Due Date: Feb 10, 2023, at 11:59 PM

Notes:

- This assignment should be finished in Java or Python.
- Zip all your files (code files, readme file, test data files, and others if needed) together into one file named “CPSC473_a#_yourlastname_yourfirstname_student#.zip” (or .rar, or .7z, or .tar), then submit this package to the Moodle.
 - For grad students, the course number is CPSC673.
- The format of input data will be discussed in class.

Assignment Description:

The main task of this assignment is straightforward: implement an Apriori algorithm for frequent pattern mining.

Detail requirements for Apriori:

1. Your program should be able to read a transaction database from a .txt file.
2. The format of the transaction database file will be discussed in class.
3. All possible input variables should be entered by command-line arguments.
 - a. For example, if, 1) the assignment is implemented in Java, 2) the executable is named “apriori”, 3) the file path of database is “./testdata.txt”, 4) the minimum support threshold is set to 50%.

The command to run your program should be:

```
java apriori ./testdata.txt 50
```

4. You should provide a README file for program compilation and execution.
5. For the output:
 - a. The first console output prints out the total number of frequent patterns.
 - i. E.g.:
 $|FPs| = 9$
 - b. The second console output prints out the execution time of your program.
 - i. E.g.:
Total Runtime: 1.445 sec
 - c. Then, output all the discovered frequent patterns into a .txt file named “MiningResult.txt”.
 - i. An example of the output file is provided as “SampleOutput.txt”. It is the mining result of the small dataset that we used in class.

Bonus Marks:

- There is a 10% bonus mark.
- To get the bonus marks for this assignment, you need to document any special enhancement you included in the implementation. For example, you used some special implementation tricks to deal with candidate generation, which, can improve the performance compared to the traditional method.
- The document does not have to be long. As long as you clearly explain a) what you did, and b) why you think this method can improve the performance.
- The document does not have to be formal. An informal report should be good enough.
- The more special enhancements you include in your implementation the better chance you will get full bonus marks.
- The document should be in PDF, and named as
“BonusMarkDocumentation_lastname_firstname.pdf”.