Homework 1: Frequent Itemset Mining

Introduction

This assignment aims to deepen your understanding of frequent itemset mining, a topic briefly covered during our lectures. You are encouraged to engage with the material more thoroughly through this report, using the provided sample reports as a guide. This task will involve studying, analyzing, implementing, and experimenting with the specified algorithms.

Instructions and Contact

Please consult the assignment manual carefully for detailed instructions. Should you have any questions, do not hesitate to reach out to the instructor, Junghoon Kim, at junghoon.kim@unist.ac.kr. For HW #1, I am going to organize a session to give feedback for each group. After that feedback, please contact me.

Topics and References

Topic 1: Apriori Algorithm

Reference: "Fast algorithms for mining association rules in large databases."

Topic 2: FP-growth Algorithm

Reference: "Mining frequent patterns without candidate generation."

~~**Topic 3: Eclat Algorithm**~~

~~Reference: "Scalable algorithms for association mining."~~

Description

We have briefly discussed frequent itemset mining in class. This report is an opportunity for a more in-depth understanding of the topic. You are expected to study, analyze, implement, and experiment with the given algorithms more thoroughly.

Implementation and Data

Sample data can be downloaded from the link provided. The data is in a CSV file format, with each line representing one transaction. Your implementation should accept the input file name and a minimum support value as input parameters. The program should operate as shown in the terminal example below: We expect that the output is sorted based on their support value. The output should be organised in ascending order according to their support values.

>java A1_G1_t1 ./groceries.csv 0.15

soda 0.1743772

rolls/buns 0.1839349

other vegetables 0.1934926

whole milk 0.255516

Additional Information

For Homework 1, we will organize a discussion session to address any difficulties encountered while working on the assignment. This session is scheduled for 04-05 and attendance is optional. It provides an opportunity to receive feedback from the instructor and fellow students. To ensure meaningful feedback, you are strongly encouraged to prepare a draft of your report in advance.

Deadline: 2024-04-12 23:59