Concordia University
Department of Computer Science and Software Engineering
COMP 6521 - Advanced Database Technology and Applications.
Fall 2018

Project 2: Frequent Item-Set Mining

Due date: Wednesday, November 14, by 15:45 p.m. EST

The task is to implement an algorithm for mining frequent pairs from market basket data. The input to the algorithm will be n baskets, each containing m items, where $3 \leq m \leq 100$. Each item is a natural number from the set $\{0, 1, 2, \ldots, 99\}$. The support threshold t, where $t \geq 3$, is also part of the input.

Your algorithm should output all frequent itemsets (not just all pairs) $\{i_1, i_2, \ldots, i_k\} \subseteq \{0, 1, \ldots, 99\}^k$ for $2 \leq k \leq 100$ of items whose support is at least t. Your algorithm should read the input as `Enter` delimited text files, where each basket is given as $basket\#, item_1, item_2, \ldots, item_m$. Every line in the Input file indicates a basket of items. $basket\#$ indicates the basket number and basket numbers are positive integers. Implement a frequent Item Set Mining Algorithm, from the literature.

The application will then output the frequent itemsets in each line, in the form

$$\{i_1, i_2, \ldots, i_k\} - Support$$

where "Support" indicates the support of itemset $\{i_1, i_2, \ldots, i_k\}$. The itemsets in teh output file should satisfy the support threshold condition.

The project will be graded based on correctness of result, wall-clock execution time on the machines in the labs, correctness and optimization of algorithm.

There are a few considerations:

- Programming language is restricted to Java.
- Your program should work with any input size.
- Execution time is important. All of the projects will be run on the machines in the lab and your grade will depend on the correctness as well as the execution time.
- You need to provide two pages of documentation describing your algorithm, optimization and contribution of each team member.
- Projects are to be submitted electronically by one of the group members in a single archive file (*.tar.gz), using the link:

  `https://fis.encs.concordia.ca/eas/`

Notice: While discussion of the project problems among students is encouraged, each group must implement the project independently. Groups should be aware of the Universitys Code of Conduct concerning cheating, plagiarism, and the possible consequences of violating this code.
Furthermore, on the submitted project you must write the following statement: "We certify that this submission is our original work and meets the Facultys Expectations of Originality", together with your signatures.