Homework Two: Mining Association Rules from Gene Expression Data
CSE601, Fall 2015

Due: Oct. 3 (hard copy before class starts, online code Oct. 2 midnight)

Problem

1. Implement the Apriori algorithm to find all frequent itemsets. Report the number of frequent itemsets for support of 30%, 40%, 50%, 60%, and 70%, respectively.
2. Generate association rules based on the templates you specify. Test templates:

- *Template 1*:
  {RULE|BODY|HEAD} HAS ({ANY|NUMBER|NONE}) OF (ITEM1, ITEM2, ..., ITEMn)
- *Template 2*:
  SizeOf({BODY|HEAD|RULE}) $\geq$ NUMBER.
- *Template 3*: Any combined templates using AND or OR. For example:
  HEAD HAS (1) OF (Disease) AND BODY HAS (NONE) OF (Disease)

The test data set is at /projects/azhang/cse601/association-rule-test-data.txt. Although the test data are gene expression data, your design should be general for all situations. The code to be submitted should be stand alone, but you may want to include this function in your project 1.

Submitting Materials

Hard copy: A comprehensive report including results should be submitted.
Online: A README file and all source codes.