CPS196.03 Information Management and Mining - Spring 2009
Assignment 2

- Due date: Thursday, Feb. 12, 2009, in class (2.50 PM). Late submissions will not be accepted.
- Submission: In class, or email solution in pdf or plain text to shivnath@cs.duke.edu.
- Do not forget to indicate your name on your submission.
- State all assumptions. For questions where descriptive solutions are required, you will be graded both on the correctness and clarity of your reasoning.
- Email questions to shivnath@cs.duke.edu.
- Total points = 100.

Question 1 **Points 10**

Describe an algorithm that can output all frequent itemsets given the set of all maximal frequent itemsets.

Question 2 **Points 15**

Describe an algorithm that can output all frequent itemsets and their supports given the set of all closed frequent itemsets and their respective supports.

Question 3 **Points 15**

Toivonen's algorithm to find frequent itemsets can stop when no itemset from the negative border is found to be frequent. Prove that this step is correct. That is, you have to show that Toivonen's algorithm will not have any false-positive or false-negative itemsets.

Question 4 **Points 15**

Question 9 on Page 407 (Chapter 6) of the reference textbook by Pang-Ning Tan and others. Chapter 6 of the reference textbook is available at: http://www-users.cs.umn.edu/~kumar/dmbook/ch6.pdf

Question 5 **Points 15**

Question 10 on Page 408 (Chapter 6) of the reference textbook by Pang-Ning Tan and others.

Question 6 **Points 15**

Question 11 on Page 409 (Chapter 6) of the reference textbook by Pang-Ning Tan and others.

Question 7 **Points 15**

Question 12 on Page 409 (Chapter 6) of the reference textbook by Pang-Ning Tan and others.