CSCI 6350 Spring 2006
Homework 3 (Reading Assignment 2 related questions)
Turn in your answers on Tuesday, Feb 7th

Read the following paper and answered the questions below.
"Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach", by J.W. Han, J. Pei, Y. W. Yin, R. Ying Mao, in *Data Mining and Knowledge Discovery*, 8, 53-87, 2004.

Questions:

1. The FP-tree takes exactly 2 scans of the transaction database, what is the purpose of each of these two scans ?
2. Show that the FP-tree is a complete representation of the transaction database.
3. Is the FP-tree the most compact representation for any transaction database? Justify your answer.
4. The FP-tree is a compact representation of the transaction database. Explain the two main steps that leads to the compactness of the tree.
5. What is the purpose of including *node links* in each of the nodes in the FP tree?
6. What is the conditional pattern base of an item a_i? Show the formal definition.
7. Suppose the support count for {I4} is 8. In item I4's conditional pattern base, {I2 I5} has support count 5. What is the support for set {I2, I4, I5}?
8. In Apriori algorithm, each iteration through the process, a set of frequent item sets of size k is derived. The size of k grows incrementally. In FP-growth method, a divide and conquer approach is used. Use the top level division as an example, how is the problem divided? What is conquered/solved at the first level?
9. What is parallel projection? What is the main problem with parallel projection?
10. What is partition projection?