Lab Work 1

(a) Implement the Apriori Algorithm. You should use a trie (or similar) data structure (or any efficient technique) while performing support-count of candidates of a particular level. Database cannot be scanner more than once for a particular level of candidate generation.
(b) Implement the FP-Growth Algorithm.
(c) Compare the performances of the Apriori and FP-Growth algorithms in real-life datasets with respect to runtime and memory. Use at least 5 datasets (more preferable) including 2 dense, 2 sparse and 1 large (Kosarak/Accidents/Webdocs etc.) datasets.
(d) Submit the source codes of (a) and (b) and a report for (c). You may use your preferred language for implementation and tool/package to prepare the comparison graphs and description of the report.

Datasets: Download real-life datasets from the following website

Frequent Itemset Mining Dataset Repository

http://fimi.uantwerpen.be/data/

Papers to get help: You may get help/idea about dataset distributions, thresholds etc. from the FP-Growth and CP-tree journal papers. Read the Fast Apriori Implementation paper to get idea about the trie-based support counting technique.