ABSTRACT

The main problem in data mining is the discovery of association rules that identify relationships among sets of items. There are two steps in mining association rules: finding the frequent itemsets and generating association rules from them. Since the mining of frequent itemsets is computationally expensive, most of the research attention has been focused on it.

Among frequent pattern discovery algorithms, FP-Growth employs search strategy using compact structure resulting in a high performance algorithm that requires only two database passes. Hereinafter, further improvements can be made to the pattern-growth approach for better performance of the mining process. The FP-Growth* is a novel array-based technique that reduces the need to traverse FP-tree. This final project implement and present experimental result of these two algorithm build in Borland Delphi 7.

The result of experiment shows that the FP-growth* algorithm run faster than FP-Growth especially for sparse datasets. Due to the use of array technique, FP-Growth* uses more memory than FP-Growth but far smaller than the size of FP-tree.

Key word: Data mining, frequent itemset, FP-tree, FP-Growth, FP-Growth*.