ABSTRACT

Scalable data mining in large databases is one of today's real challenges to database research area. The integration of data mining with database systems is an essential component for any successful large scale data mining application.

A fundamental component in data mining tasks is finding frequent patterns in a given dataset. Most of the previous studies adopt an Apriori like candidate set generation and test approach. However, candidate set generation is still costly, especially when the database is large.

This final project implement and present experimental result of SQL based frequent pattern mining with a novel frequent pattern growth (FP-growth) method, which is efficient and scalable for mining frequent patterns without candidate generation.

Key word: Data mining, association rule, SQL based frequent pattern mining, frequent pattern-growth (FP-growth).