ACKNOWLEDGEMENT

This research work is devoted to the OMNIPRESENT, ALMIGHTY for ever blessing in my endeavors and showering to my father, mother and my husband.

My beloved thanks to Dr.K.Manimekalai, Vice-Chancellor, Mother Teresa Women’s University, Kodaikanal, The Registrar, Mother Teresa Women’s University, Kodaikanal, Dr.Rathidevi, Dean (Research), Mother Teresa Women’s University, Kodaikanal and Dr.Pushparani, Head of the Department, Computer Science, Mother Teresa Women’s University, Kodaikanal for providing me an opportunity to complete and submit this thesis work.

I express my heartfelt gratitude to my research supervisor Dr.R.Balasubramanian for his precious guidance, innovative thinking and keen interest in shaping this thesis to its perfection. I thank the almighty for giving me such a wonderful person to foster my succession.

I wish to extend my thanks to all staff members, Department of Computer Science, Mother Teresa Women’s University, Kodaikanal for their constant support and encouragement.

Eventually, it is needless to say that my gratitude goes towards all my friends who supported me in this adventure.

Although they knew nothing about it, my daughter G. Indupriyaa and my son G. Vijay Krishnan have been the nicest possible distraction.

Finally, I would like to convey my dedication to my family member’s, father-in-law and mother-in-law for their patience, sacrifice and all the encouragement without which it would not have been possible for me to complete my study successfully.
Most algorithm works on memory data structure or from flat-file. Flat-files are disk based files which reduces the computational cost of algorithms. Lot of algorithms are developed in the field of association rule mining based on the existing fundamental algorithms such as Apriori, FP-Growth and Eclat etc.

Searching methods use the index structures to gain fast access to records in a file. The indexes are mainly used to reduce the scanning time of the entire data base. Each index structure is associated with a particular search key. An index stores the values of the search key in sorted order, and associates with each search key the records that contain that search key. For some reasons the searching algorithms are seldom applied on relational databases which store and manage very large volumes of data. The main reason is the relational data bases have the power and speed to traverse large data within a short span of time.

The proposed Refined Search Divide and Conquer (Hybrid) Algorithm (RSDCA) is applied to search frequent itemset mining on very large databases. This research proposes RSDCA to exhibits diminishing on the time factor by using the indexes of the databases and also examines a generic way of identifying the indexes in the databases using divide and conquer methods.

The earliest algorithms and researchers of different study were used their own memory data structure for retrieving datasets and executing their algorithms. These structures inflict a limitation on the size of data that can be processed. But RDBMS allow its user/applicant to utilize its benefits of using its buffer management systems. So the applicant can be free from size consideration of the data.
The proposed RSDCA explores new possibilities on the generic searches executed on Very Large Data Bases (VLDB). This algorithm needs proper indexing and more particularly generic indexing to improve the speed of queries.

The proposed algorithm used divide and conquer method for mining frequent pattern. The rapid searches itemset mining algorithm rapidly reads transactions and updates support counts at the same time. The Divide and Conquer method is powerful tool for solving difficult problems. The divide and conquer method divide the problem into several sub problems. The dividing part picks which segment in which to search, and combine the solutions. This similar principle is applied in several data structure, which is also applied in the proposed RSDCA to extract the frequent itemsets.

Divide and conquer algorithms tend to make efficient use of memory cache. Though it divide the problems into sub problems, the sub problems are small enough and be solved within the cache, without accessing the main memory.

In this work, a class of novel and efficient pattern-growth methods for mining various frequent patterns from large databases is developed. Pattern growth methods adopt a divide-and-conquer approach to decompose both the mining tasks and the databases and avoid the costly candidate-generation-and-test processing completely.

The Refined Search Divide and Conquer (Hybrid) Algorithm (RSDCA) is compared with various existing algorithm such as Apriori, FP-Growth, and Eclat. The execution time of the proposed algorithms is very less than the above mentioned algorithms for different dataset.