

Задание 1. Поиск частых наборов

Выполните поиск частых наборов объектов в трех различных наборах данных с помощью следующих алгоритмов (или их модификаций): Apriori, FP-Growth, ECLAT. Наборы данных должны существенно отличаться друг от друга по количеству транзакций и/или типичной длине транзакции (количеству объектов). Варьируйте пороговое значение поддержки (например: 1%, 3%, 5%, 10%, 15%, 20%). Проверьте идентичность результатов, полученных с помощью различных алгоритмов.

1. Подготовьте список частых наборов, в которых не более семи объектов (разумное количество). Проанализируйте и изложите содержательный смысл полученного результата.
2. Выполните визуализацию полученных результатов в виде следующих диаграмм:
 - сравнение быстродействия алгоритмов на фиксированном наборе данных при изменяемом пороге поддержки;
 - общее количество частых наборов объектов на фиксированном наборе данных при изменяемом пороге поддержки;
 - максимальная длина частого набора объектов на фиксированном наборе данных при изменяемом пороге поддержки;
 - количество частых наборов объектов различной длины на фиксированном наборе данных при изменяемом пороге поддержки.

Типовые контрольные вопросы

1. Дайте определения следующих понятий: база транзакций, поддержка, частый набор.
2. Объясните принцип антимонотонности поддержки.
3. Объясните работу одного из следующих алгоритмов поиска частых наборов: Apriori, FP Growth, ECLAT.