

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное
учреждение высшего образования
**"Южно-Уральский государственный университет
(национальный исследовательский университет)"**
Высшая школа электроники и компьютерных наук
Кафедра системного программирования

ОТЧЕТ
по практической работе 1
«Поиск частых наборов»
по дисциплине
«Технологии аналитической обработки информации»

Выполнил: _____
студент группы КЭ-403
О.С. Мазжухин

Проверил: _____
преподаватель
А.И. Гоглачев
Дата: _____
Оценка: _____

Формулировка задания

1. Разработайте программу, которая выполняет поиск частых наборов объектов в заданном наборе данных с помощью алгоритма Apriori (или одной из его модификаций). Список результирующих наборов должен содержать как наборы, так и значение поддержки для каждого набора. Параметрами программы являются набор, порог поддержки и способ упорядочивания результирующего списка наборов (по убыванию значения поддержки или лексикографическое).

2. Проведите эксперименты на наборе данных `baskets.csv` (сведения о покупках в супермаркете). В экспериментах варьируйте пороговое значение поддержки (например: 1%, 3%, 5%, 10%, 15%).

3. Выполните визуализацию результатов экспериментов в виде следующих диаграмм:

- сравнение быстродействия на фиксированном наборе данных при изменяемом пороге поддержки;
- количество частых наборов объектов различной длины на фиксированном наборе данных при изменяемом пороге поддержки.

Гиперссылка на каталог репозитория с исходными текстами, наборами данных и другими материалами:

<https://github.com/LN4rkot1k/informationProcessing>

Визуализация

Сравнение быстродействия на фиксированном наборе данных при изменяемом пороге поддержки представлено на рисунке 1.

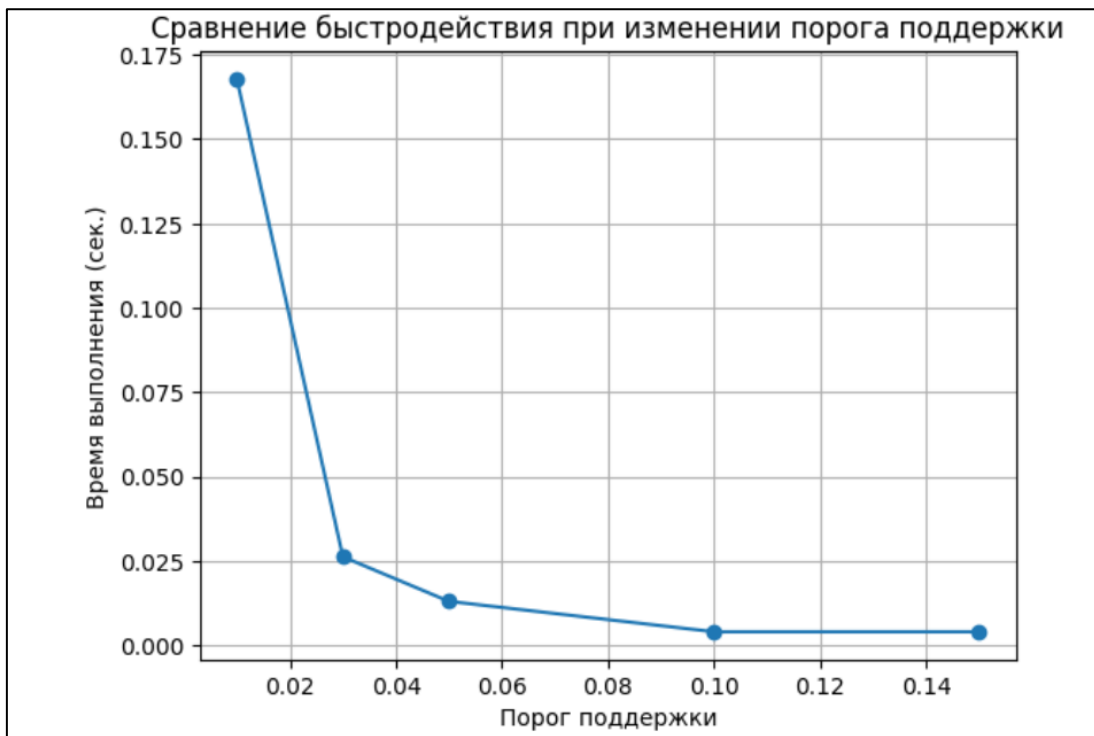


Рисунок 1 – График зависимости времени от порога поддержки

Глядя на график, можно заметить следующую зависимость: чем выше порог поддержки, тем меньше время выполнения, потому что при высоких порогах алгоритм находит меньше частых наборов, соответственно, выполняется быстрее. На малых порогах алгоритм обрабатывает очень много комбинаций товаров, но при увеличении порога многие редкие комбинации отбрасываются, снижая вычислительную нагрузку.

Количество частых наборов объектов различной длины на фиксированном наборе данных при изменяемом пороге поддержки представлено на рисунке 2.

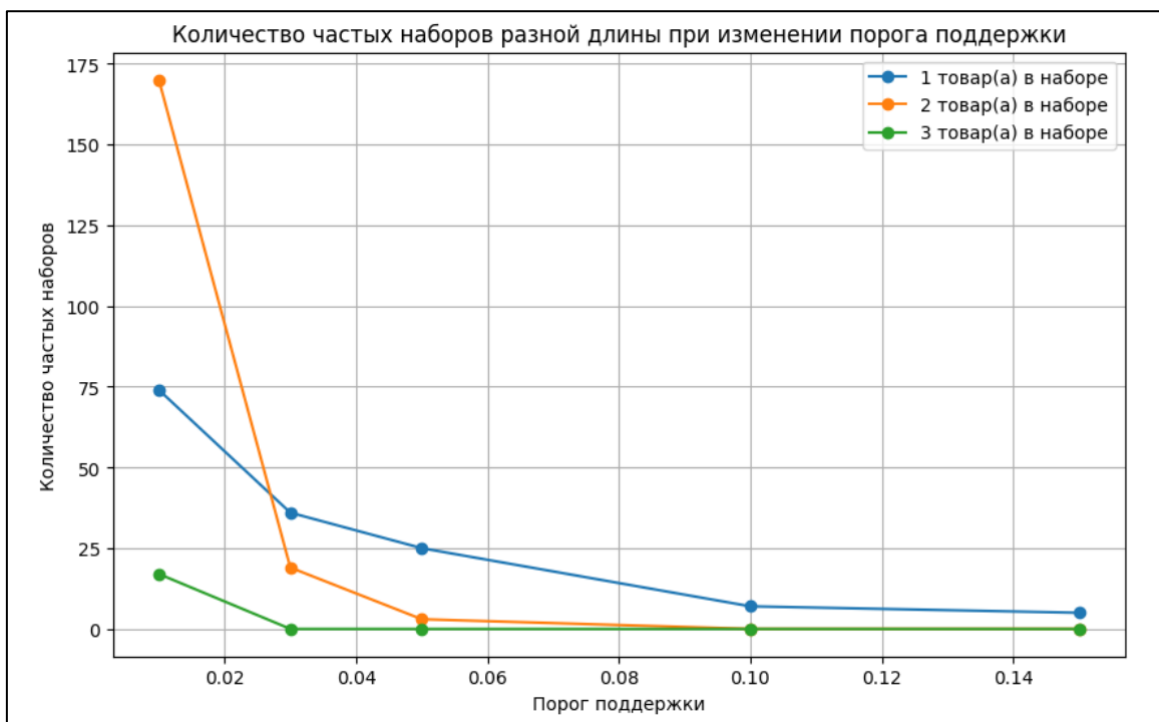


Рисунок 2 – График зависимости количества частых наборов разной длины от изменения порога поддержки

Все три линии сначала резко падают, а затем стабилизируются, потому что при низком пороге много редких комбинаций попадает в частые наборы, но при увеличении порога редкие комбинации отсеиваются.