

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-404  
А.Ю. Емельянова

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:

1) сравнение быстродействия на фиксированном наборе данных при изменяемом пороге поддержки;

2) 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/Sunless-sea/Technologies-of-analytical-information-processing>

## **Визуализация**

В данной практической работе были проведены некоторые эксперименты при изменении порога поддержки для частых наборов товаров. Также были выявлены некоторые зависимости между временем обработки данных, количеством наборов и различными порогами поддержки.

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

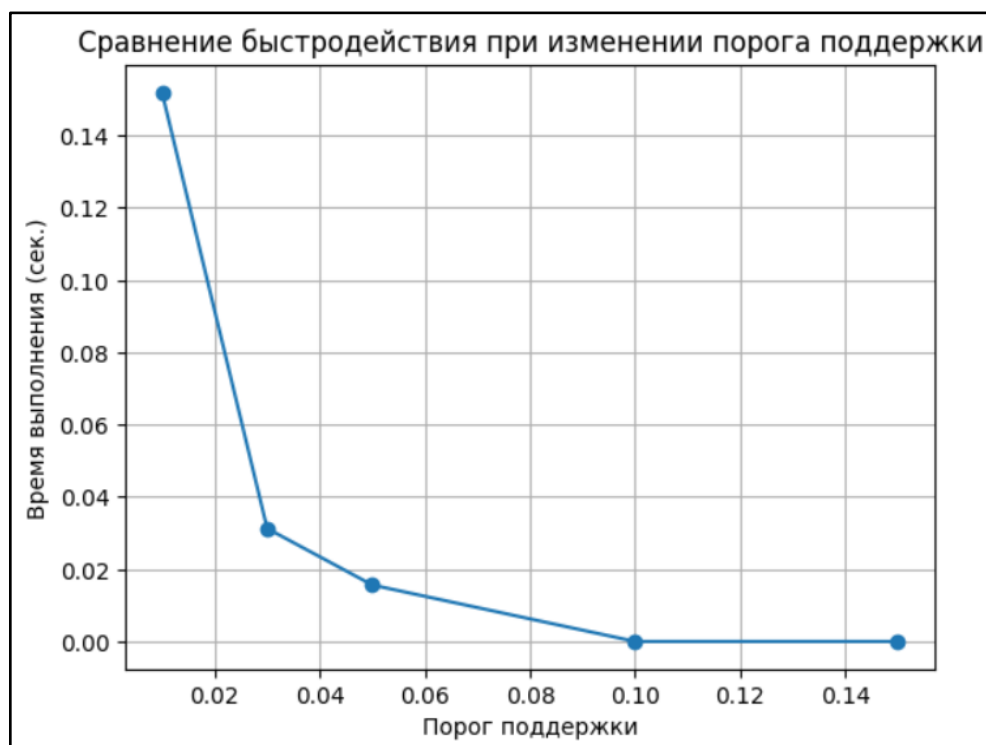


Рисунок 1 – График зависимости времени от порога поддержки

Из анализа графика можно сделать следующий вывод: с увеличением порога поддержки время выполнения сокращается. Это объясняется тем, при повышении порога поддержки время выполнения уменьшается, так как при высоких порогах поддержки алгоритм находит меньше частых наборов, а значит выполняется быстрее. Напротив, при низких порогах поддержки алгоритму приходится обрабатывать большое количество возможных наборов товаров, но по мере повышения порога редкие наборы отбрасываются, снижая вычислительную нагрузку.

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

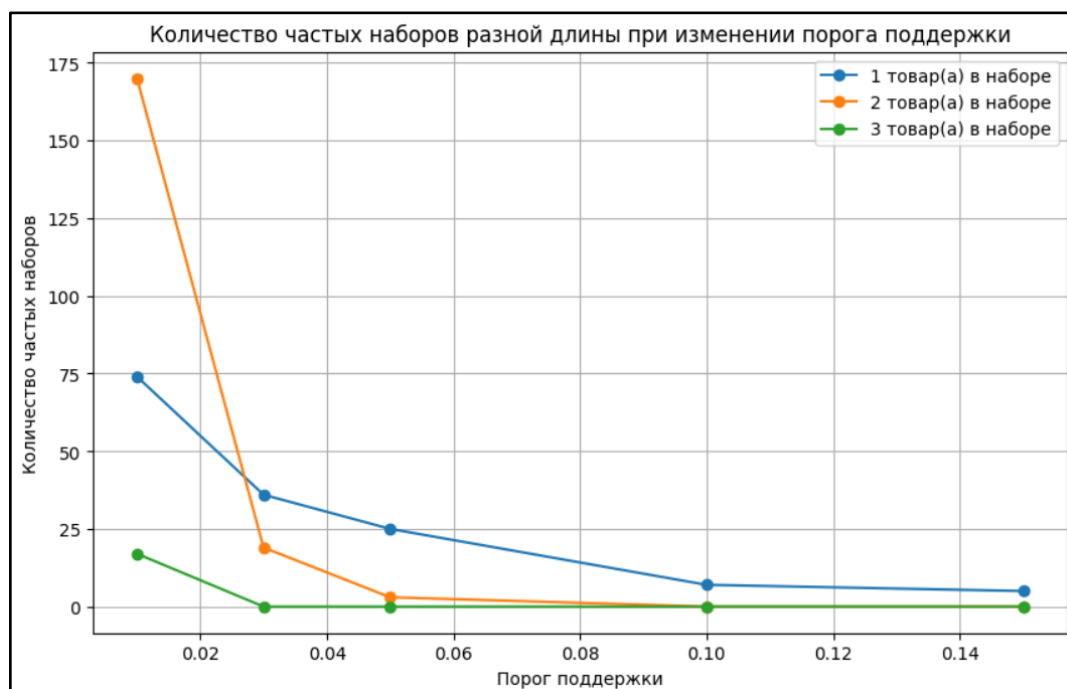


Рисунок 2 – График зависимости количества частых наборов разной длины от изменения порога поддержки

Количество частых наборов резко уменьшается при увеличении порога поддержки на всех выборках, так как порог поддержки отсекает маловстречающиеся наборы. Однако, затем они стабилизируются примерно на уровне порога в 10% частоты, так как остаются наиболее встречающиеся наборы.