

**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**
Санкт-Петербургская школа физико-математических и компьютерных
наук

ОТЧЁТ О ЛАБОРАТОРНОЙ РАБОТЕ
по курсу: «Алгоритмы и структуры данных»
по теме: «Сжатие данных кодом Хаффмана»

Выполнил: студент группы БКОТСИС-241
Феськов Герман Дмитриевич
Преподаватель: к.т.н., доцент А.А. Овчинников

Санкт-Петербург, 2025

1 Постановка задачи

Реализовать сжатие произвольного файла кодом Хаффмана с алфавитом из 256 символов. Использовать запись преамбулы в виде битового представления бинарного дерева. Исследовать избыточность кода Хаффмана: провести не менее 100 экспериментов для p_1 от 0.1 до 0.9 с шагом 0.1. Сравнить с границей Галлагера.

2 Алгоритм Хаффмана

Метод сжатия данных, основанный на использовании префиксного кода, минимизирующего среднюю длину кодового слова. Дерево Хаффмана строится следующим образом:

1. Определяются вероятности появления символов в тексте.
2. Создаётся очередь с приоритетами, содержащая узлы с символами и их вероятностями.
3. Повторяются следующие шаги, пока в очереди не останется один узел:
 - Извлекаются два узла с наименьшими вероятностями.
 - Создаётся новый узел, объединяющий два узла, и его вероятность равна сумме вероятностей объединяемых узлов.
 - Новый узел добавляется обратно в очередь.
4. Символы кодируются в зависимости от пути в дереве: 0 — левый потомок, 1 — правый потомок.

Для построения дерева была использована `priority_queue` с собственным компаратором. Реализован Go-скрипт, а также битовый обработчик для работы с битовым представлением данных.

3 Эксперименты и анализ

Для исследования избыточности кода Хаффмана проводились эксперименты при различных распределениях вероятностей символов. Сравнение проведено с границей Галлагера, которая определяется как

$$H(X) + \frac{1}{2 \ln 2} \sum_i p_i \log p_i. \quad (1)$$

Результаты экспериментов представлены в таблицах и графиках (Приложение А).

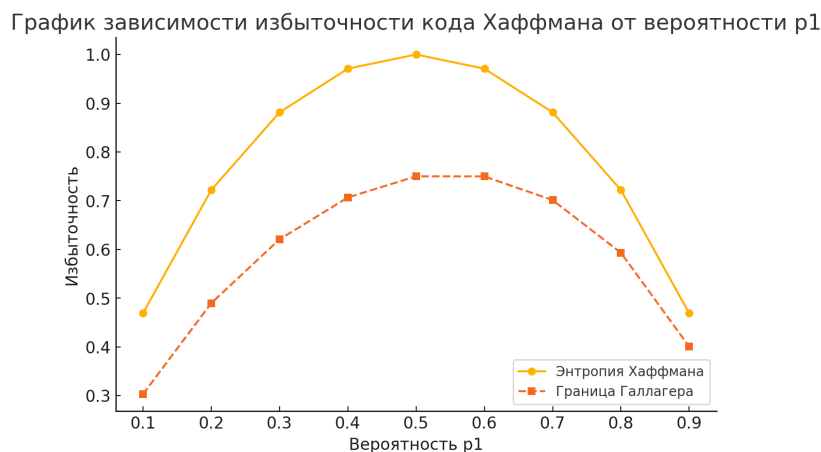


Рис. 1: График зависимости избыточности кода Хаффмана от вероятности p_1

4 Заключение

Была реализована программа для сжатия данных кодом Хаффмана. При этом использованы:

- очередь с приоритетами `priority_queue` с пользовательским компаратором для построения дерева Хаффмана;
- скрипт на Go для работы с кодированием и декодированием данных;
- битовый обработчик для эффективного хранения и обработки закодированных данных.

В результате экспериментов были получены оценки избыточности кода Хаффмана при различных вероятностных распределениях. Сравнение с границей Галлагера показало, что на практике код Хаффмана демонстрирует близкие к теоретическим значения эффективности сжатия, но в некоторых случаях присутствует небольшая избыточность. Результаты экспериментов подтверждают соответствие теоретическим ожиданиям, что делает метод кодирования Хаффмана эффективным средством для сжатия данных в широком спектре приложений.

Исходный код

1. Репозиторий проекта на GitHub

Список использованных источников

1. Кудряшов Б.Д. «Теория информации».
2. Хаффман Д. «Методы кодирования информации».