Учреждение образования «БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ  
ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Факультет информационных технологий

[Кафедра информационных](https://www.belstu.by/fakultety/fit/vm) систем и технологий

Специальность 1-40 01 01 Программное обеспечение информационных технологий

Лабораторная работа №2

**Элементы теории информации. Параметры и характеристики дискретных информационных систем**

Студент: Корнелюк. В. В.

ФИТ 3 курс 4 группа

Преподаватель: Нистюк О.А.

Минск, 2025

**Лабораторная работа №2**

**Цель:** приобретение практических навыков расчета и анализа параметров и информативных характеристик дискретных ИС.

**Задачи:**

1. Закрепить теоретические знания по основам теории информации.

2. Разработать приложение для расчета и анализа параметров и информативных характеристик дискретных ИС.

3. Результаты выполнения лабораторной работы оформить в виде описания разработанного приложения, методики выполнения экспериментов с использованием приложения и результатов эксперимента.

# Теоретические сведения

Передача информации (данных) осуществляется между двумя абонентами, называемыми источником сообщения (ИcС) и получателем сообщения (ПС). Третьим элементом информационной системы является канал (среда) передачи, связывающий ИсС и ПС.

ИсС и ПС обмениваются информацией в технических системах в виде сигналов, сформированных на основе определенного алфавита. Характеристикой алфавита является его мощность, N – количество символов, на основе которых формируется сообщение. Например, мощность английского алфавита – 26 символов, русского – 33 символа, мощность алфавита, на основе которого функционируют и взаимодействуют между собой компьютеры, составляет 2 символа (0 и 1).

В произвольном сообщении символы алфавита могут появляться с различной вероятностью. Если длина сообщения достаточно велика, то статистический анализ этого сообщения позволит получить вероятностные характеристики данного алфавита. Очевидно, что отличные символы в произвольном сообщении (особенно при N> 2) появляются с различной вероятностью.

Информационной характеристикой алфавита (источника сообщений на основе этого алфавита) является энтропия.

Этот термин применительно к техническим системам был введен К. Шенноном и Р. Хартли.

Энтропию алфавита А{} по Шеннону рассчитывают по следующей формуле:

где P() – вероятность P(ξ=); – элемент алфавита,.

С физической точки зрения энтропия показывает, какое количество информации приходится в среднем на один символ алфавита. Частным случаем энтропии Шеннона является энтропия Хартли. Дополнительным условием при этом является то, что все вероятности одинаковы и постоянны для всех символов алфавита.

С учетом этого формулу энтропии Шеннона можно преобразовать к виду:

Сообщение M, которое состоит из n символов, должно характеризоваться определенным количеством информации I(M):

Здесь Н(А) – энтропия алфавита с соответствующим распределением вероятностей р().

Двоичный канал передачи информации является дискретным – он основан на алфавите, состоящем из двух символов: 0 и 1 – A {0,1}.

Если вероятность ошибки в ДСК отлична от 0 (р > 0), переданное сообщение может содержать ошибки: Хk ≠ Yk. Количество информации в таком сообщении при его передаче по ДСК будет определяться не энтропией двоичного алфавита, а эффективной энтропией He(A) алфавита или пропускной способностью канала:

,

где – условная энтропия:

# Ход работы

Для вычисления энтропии было разработано несколько функций.

Функция calculateSymbolsFrequency принимает имя файла и строку алфавита. Она выполняет чтение файла, подсчитывает частоту появления символов данного алфавита и общее количество прочитанных символов. Код функции представлен на рисунке 2.1.

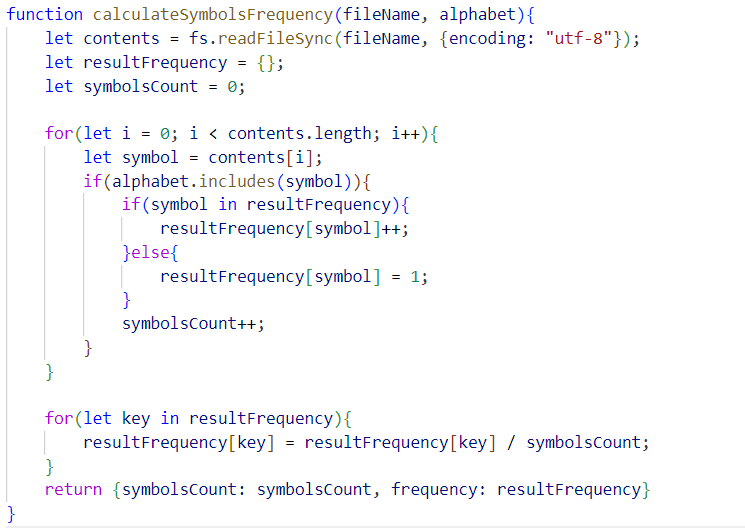


Рисунок 2.1 – Функция для подсчета частоты появления символов алфавита

На основе данных, полученных при работе данной функции, были построены гистограммы частот алфавитов. Гистограмма частот появления символов голландского алфавита представлена на рисунке 2.2.

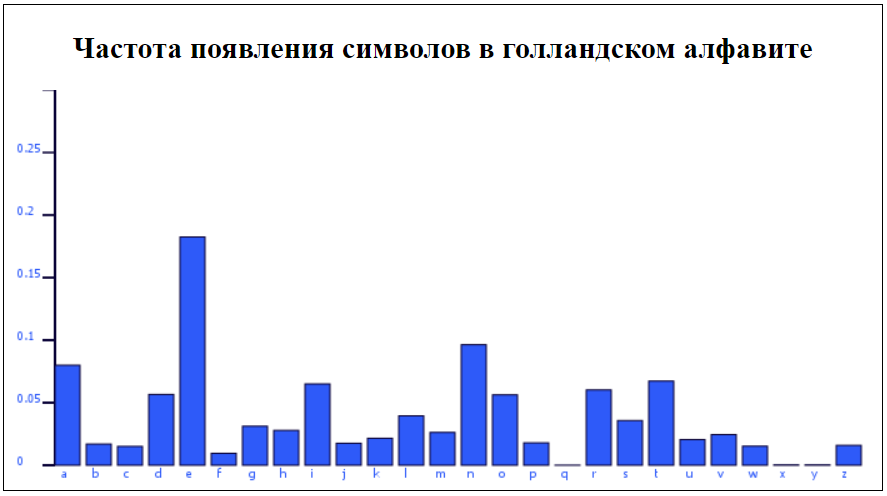


Рисунок 2.2 – Гистограмма частот для голландского алфавита

Гистограмма частот появления символов украинского алфавита представлена на рисунке 2.3.



Рисунок 2.3 – Гистограмма частот для украинского алфавита

Затем, была разработана функция calculateEntropy, которая на основе данных о частоте появления символов вычисляет энтропию алфавита с использование формулы энтропии по Шеннону. Код функции представлен на рисунке 2.4.

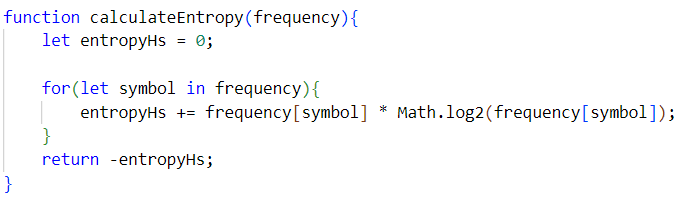


Рисунок 2.4 – Функция для подсчета энтропии алфавита по Шеннону

Результат работы функций представлен на рисунке 2.5.

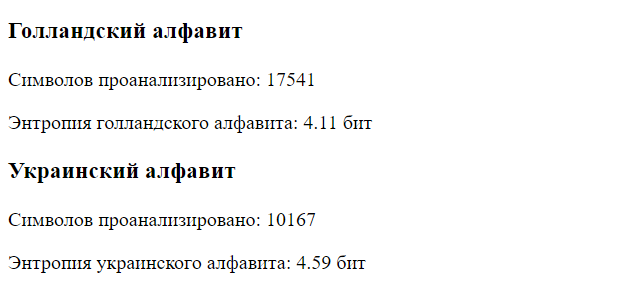


Рисунок 2.5 – Энтропия голландского и украинского алфавитов

Далее, нужно было подсчитать энтропию бинарного алфавита. Для этого, символы текста были представлены в бинарных кодах. Результат вычисления энтропии бинарного алфавита представлен на рисунке 2.6.

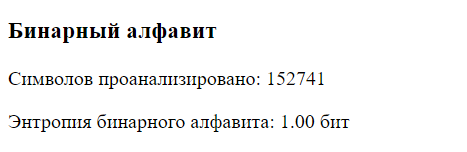


Рисунок 2.6 – Энтропия бинарного алфавита

В следующем задании было нужно подсчитать количество информации в сообщении, состоящем из собственных фамилии, имени и отчества (на основе исходного алфавита и в кодах ASCII). Для этого, была разработана функция calculateInformationCount, которая принимает сообщения и энтропию алфавита. Код функции представлен на рисунке 2.7.

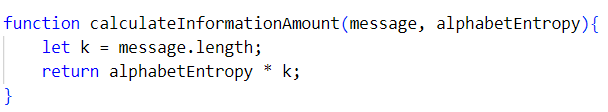


Рисунок 2.7 – Функция для подсчета количества информации в сообщении

Результат работы функции представлен на рисунке 2.8.

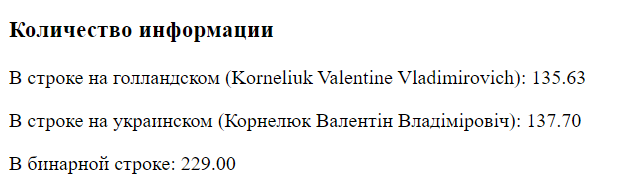


Рисунок 2.8 – Количество информации в сообщениях

Можно заметить, что количество информации в бинарной строке значительно превышает количество информации в строках на голландском и на украинском языках. Данное отличие можно объяснить отсутствием избыточности в бинарном алфавите.

В последнем задании, требовалось найти количество информации при условии, что вероятность ошибочной передачи единичного бита сообщения составляет: 0,1; 0,5; 1,0. В таком случае, количество информации в таком сообщении при его передаче по ДСК будет определяться эффективной энтропией. Для определения эффективной энтропии была разработана функция calculateEffectiveEntropy, которая принимает энтропию алфавита по Шеннону и вероятность ошибочной передачи единичного бита. Реализация функции представлена на рисунке 2.9.

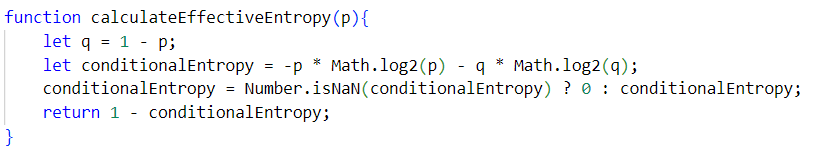


Рисунок 2.9 – Функция для определения эффективной энтропии

Результат работы функции представлен на рисунке 2.10.

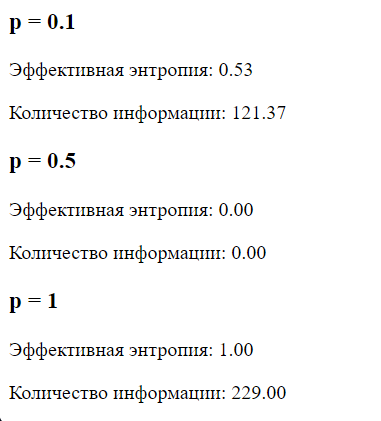


Рисунок 2.10 – Количество информации, если есть вероятность ошибки

Когда вероятность ошибочной передачи единичного бита равна 1, то количество информации будет примерно таким же, как и при вероятности ошибки равной 0, поскольку все биты исходного сообщения инвертированы, а значит при применении операции сложения по модулю 2 сообщение может быть восстановлено.

**Вывод:** в ходе выполнения данной лабораторной работы была определена энтропия голландского, украинского и бинарного алфавитов, а также было определено количество информации в сообщениях на основе этих алфавитов. Кроме того, было определено количество информации при условии, что есть вероятность ошибочной передачи единичного бита.