МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образование

«Белорусский государственный технологический университет»

Кафедра информационных систем и технологий

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

Студент:

Миневич Кристина Викторовна

Вариант 7

Минск 2023

1. **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.

Отображение сообщения обеспечивается изменением какой-либо физической величины, характеризующей процесс (например, амплитуда, частота, фаза). Эта величина является *информационным* *параметром сигнала* (в общем случае–информационной системы).

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).

Сигналы, как и сообщения, могут быть непрерывными и дискретными. Информационный параметр непрерывного сигнала с течением времени может принимать любые мгновенные значения в определенных пределах. Непрерывный сигнал часто называют аналоговым, а каналы и устройства, функционирующие на основе такого типа сигналов – аналоговыми. Дискретный сигнал (устройство или канал передачи) характеризуется конечным числом значений информационного параметра.

Дискретные сообщения состоят из последовательности дискретных знаков. Часто этот параметр принимает всего два значения (0 или 1).

Сообщение или канал его передачи на основе этих двух значений сигнала называют двоичным или бинарным. Построение сигнала по определенным правилам, обеспечивающим соответствие между сообщением и сигналом, называют кодированием.

Кодирование в широком смысле– преобразование сообщения в сигнал.

Кодирование в узком смысле – представление исходных знаков, называемых символами, в другом алфавите с меньшим числом знаков. Оно осуществляется с целью повышения надежности и преобразования сигналов к виду, удобному для передачи по каналам связи. Последний тип кодирования относится к так называемой прикладной теории кодирования информации, занимающейся поиском и реализацией методов и средств обнаружения несоответствий (ошибок) между переданным и принятым сообщениями.

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.

(1.1)

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 (1.1) можно преобразовать к виду:

(1.2)

Например, энтропия Хартли для латинского (английского) алфавита составляет 4,7 бит. Если подсчитать энтропию Шеннона и энтропию Хартли для одного и того же алфавита, то они окажутся неравными. Это несовпадение указывает на избыточность любого алфавита (при N> 2).

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

(1.3)

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

Нетрудно предположить и просто убедиться, что количество информации в сообщении, подсчитанное по Шеннону, не равно количеству информации, подсчитанному по Хартли. На основе этого парадокса строятся и функционируют все современные системы сжатия (компрессии) информации.

Любые сообщения характеризуются информационной избыточностью, что позволяет сжимать их без потери информации.

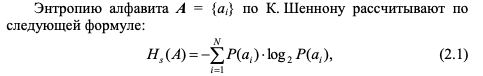
Двоичный канал передачи информации является дискретным – он основан на алфавите, состоящем из двух символов: 0 и 1 – A {0,1}.

Полагая, что сообщение М состоит только из единиц (М = 11…1) и имеет длину n: 111…11, вероятность того, что произвольный символ равен единице, составляет единицу (Р(1) = 1); другая вероятность – Р(0) = 0 для . Здесь имеет место использование моноалфавита: алфавита, состоящего из одного символа.

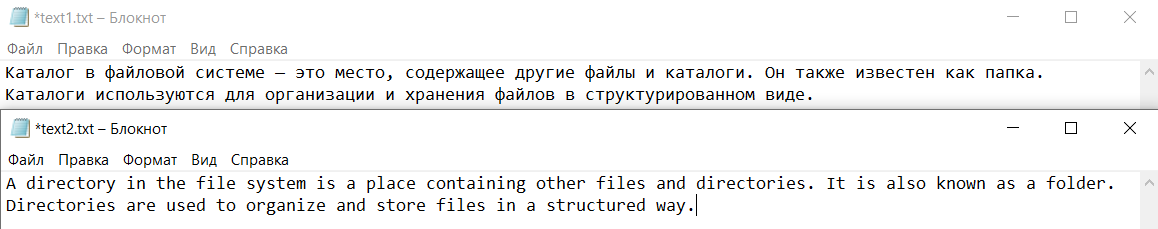
Физический смысл понятия информации в теории Шеннона: информацией является лишь такое сообщение, которое снимает некоторую неопределенность, т. е. содержит новые для получателя данные.

1. **Практическая часть**

В данной лабораторной работе необходимо было создать приложение для расчета и анализа параметров и информативных характеристик дискретных ИС**.**

**ЗАДАНИЕ 1.** Рассчитать энтропию указанного преподавателем алфавитов: один – на латинице, другой – на кириллице (по формуле (2.1) – перейти от частоты появления каждого символа алфавита к соответствующей вероятности); в качестве входного может быть принят произвольный электронный текстовый документ на основе соответствующего алфавита; частоты появления символов алфавитов оформить в виде гистограмм (можно воспользоваться приложением MS Excel);

Реализация функции расчета энтропии по К. Шеннону на языке JavaScript:

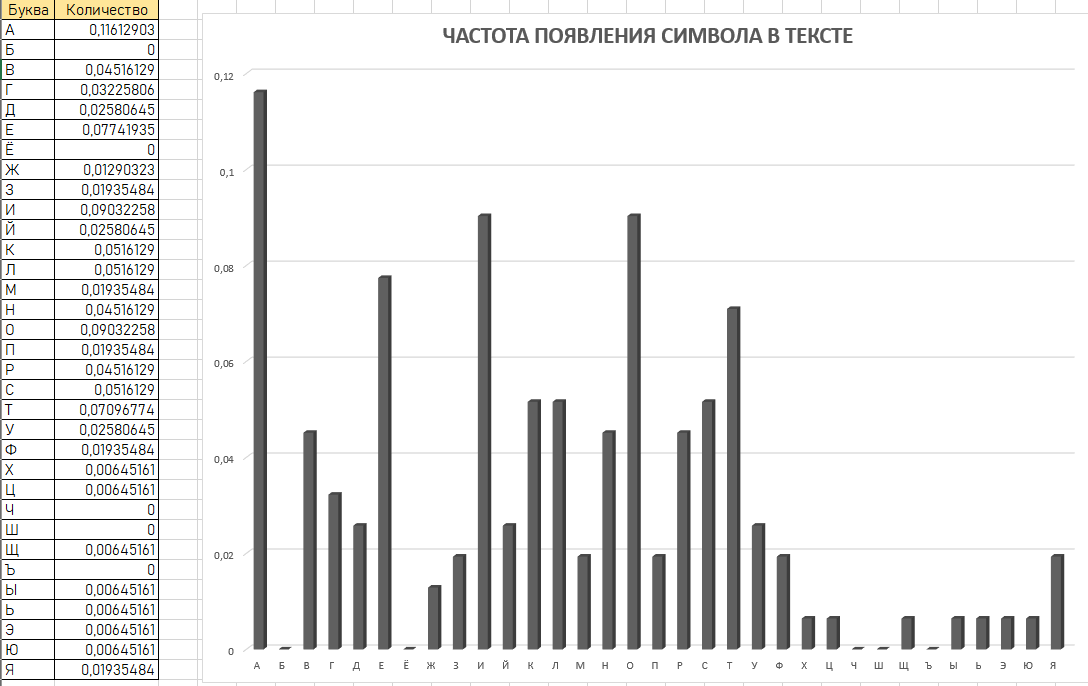
Содержимое текстовых файлов на русском и английском языках представлено ниже:  


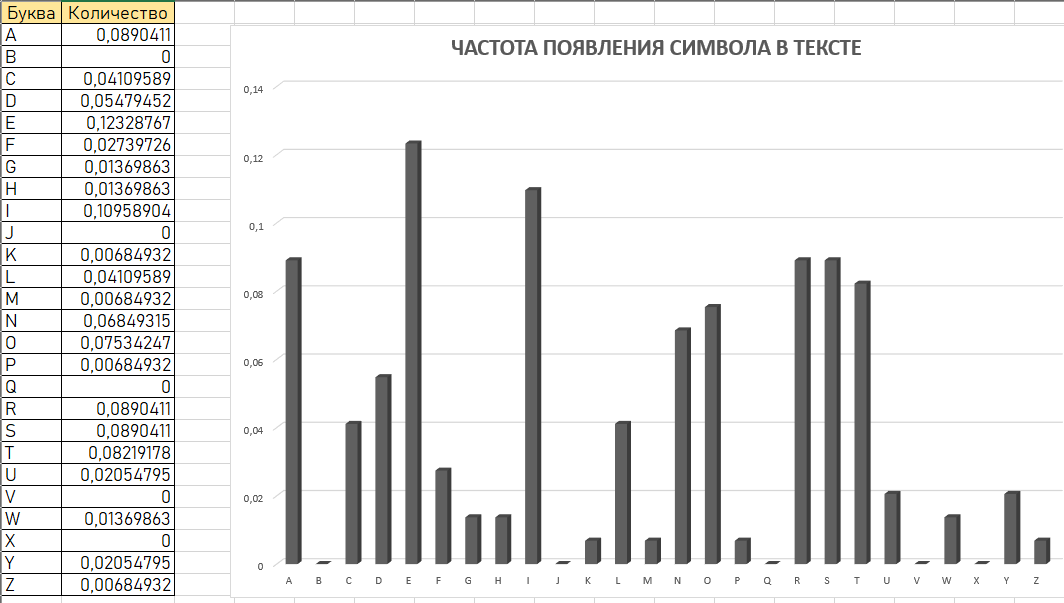
Результат работы функции подсчета энтропии на К. Шеннону для текстов на русском и английском языках:





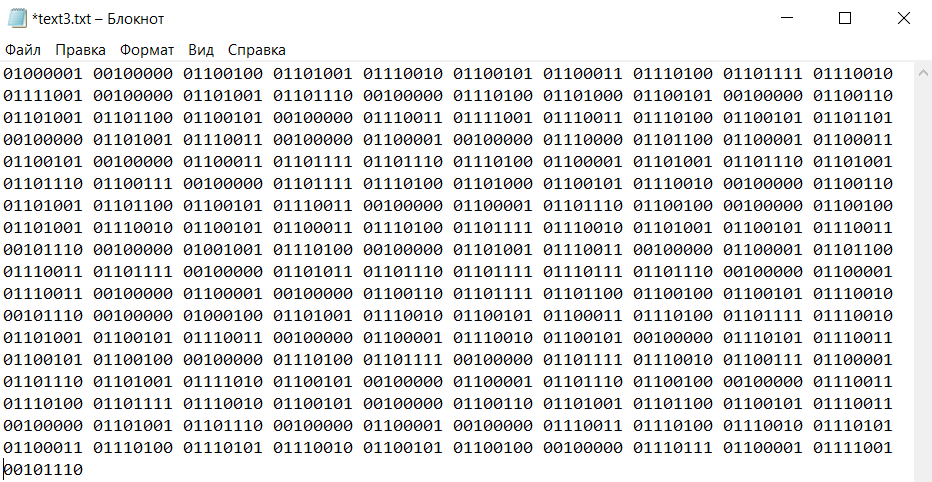
Диаграммы частоты появления символов алфавитов:

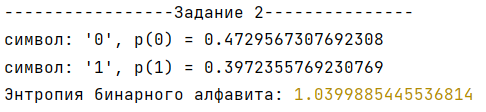




**ЗАДАНИЕ 2.** Для входных документов, представленных в бинарных кодах, определить энтропию бинарного алфавита;

Исходный текстовый документ



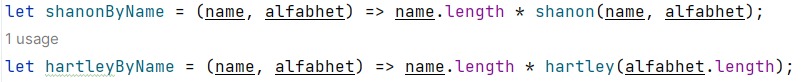
Опять используем функцию расчета энтропии по К. Шеннону, приведенную в прошлом задании. Результат работы представлен ниже:

**ЗАДАНИЕ 3.** Используя значения энтропии алфавитов, полученных в пп. а) и б), подсчитать количество информации в сообщении, состоящем из собственных фамилии, имени по отчества (на основе исходного алфвита – а) и в кодах ASCII – б); объяснить полученный результат;



Исходная строка:



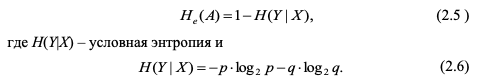
Реализация функций подсчета количества информации в сообщении:

Результат работы:

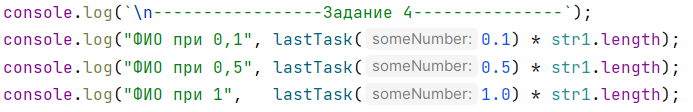
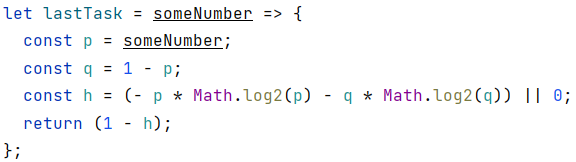


**ЗАДАНИЕ 4.** Выполнить задание п. в) при условии, что вероятность ошибочной передачи единичного бита сообщения составляет: 0.1; 0.5; 1.0.

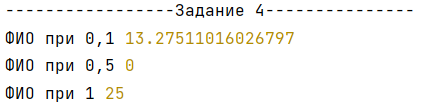
Формула расчета эффективной энтропии алфавита или пропускной способностью канала:



Реализация функции расчета на языке JavaScript:



Результат работы:



**Вывод:** таким образом, в результате лабораторной работы были получены знания по нахождению энтропии алфавита и информативности сообщения. Так же был изучен способ нахождения информативности, при различных вероятностях ошибки во время передачи, сообщения.