Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

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

Студент: Шулаков Андрей

ФИТ 3 курс 5 группа

Преподаватель: Савельева М.Г.

Минск 2023

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

**Задачи**:

* Закрепить теоретические знания по основам теории информации;
* Разработать приложение для расчета и анализа параметров и информативных характеристик дискретных ИС;
* Результаты выполнения лабораторной работы оформить в виде описания разработанного приложения, методики выполнения экспериментов с использованием приложения и результатов эксперимента.

# **Теоретические сведения**

*Основные понятия из предметной области.*

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.

Отметим также, что и в системах с хранением информации всегда можно выделить ИcС и ПС. В данном случае каналом передачи здесь выступает устройство хранения информации (память). Например, при записи данных в ОЗУ (оперативное запоминающее устройство) компьютера в качестве ИcС и ПС может выступать процессор (соответственно при записи и чтении данных).

Таким образом, простейшая информационная система состоит из трех элементов: источника сообщения, канала передачи сообщения и получателя сообщения.

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).

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 *Xk* и принятым *Yk* сообщениями.

*Понятия и формулы, использованные в лабораторной работе.*

**Алфавит, А**– это общее число знаков или символов (*N*), используемых генерации или передачи сообщений. Символы алфавита будем обозначать через *{аi}*, где *1 ≤ i≤ N*; *N* – мощность алфавита.

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

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

Энтропию алфавита *А={ai}* по К. Шеннону рассчитывают по следующей формуле:

(2.1)

С физической точки зрения энтропия алфавита показывает, какое количество информации (бит) 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 можно преобразовать к виду:

(2.2)

Сообщение *Хk*, которое состоит из *k* символов, должно характеризоваться определенным количеством информации, *I(Хk)*:

(2.3)

Для ДСК: *p + q = 1*, где *q* – вероятность правильной передачи бита сообщения, а *p* – вероятность передачи бита с ошибкой.

(2.4)

где *H(Y|X)* – условная энтропия

(2.5)

1. **Описание разработанного приложения**

Приложение написано на языке программирования C# и позволяет:

* рассчитать энтропию португальского и болгарского алфавитов; в качестве входного принят электронный текстовый документ;
* сгенерировать гистограммы частот появления символов (MSExcel);
* определить энтропию бинарного алфавита для входного документа;
* подсчитать количество информации в сообщении, состоящем из ФИО;
* выполнить предыдущее задание при условии, что вероятность ошибочной передачи единичного бита сообщения составляет: 0.1, 0.5, 1.0 .

1. **Методика выполнения поставленных задач**

Для расчета энтропии португальского и болгарского алфавита с помощью разработанного нами приложения текст из входного файла "portugal.txt" и "bulgarian.txt" записывается в строку, которая затем передается в качестве параметра при вызове функции, рассчитывающей энтропию алфавита.

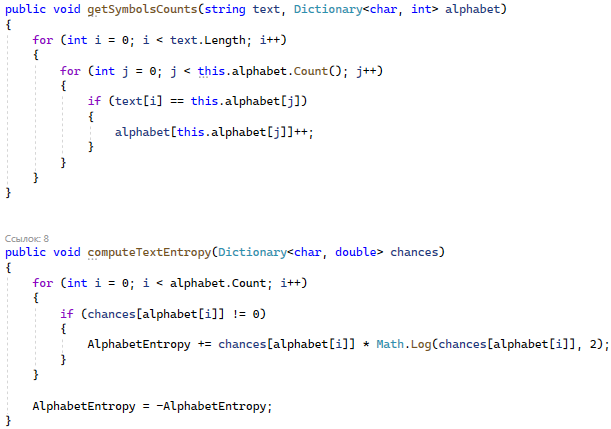


Рисунок 1 – описание функций для расчета энтропии алфавита

В результате выполнения данной функции количество вхождений и вероятность появления каждого символа показывается в консоли приложения, а также возвращает значение энтропии, рассчитанное по формуле Шеннона: .

Помимо этого приложение рассчитывает количество информации в данном входном файле по формуле: .

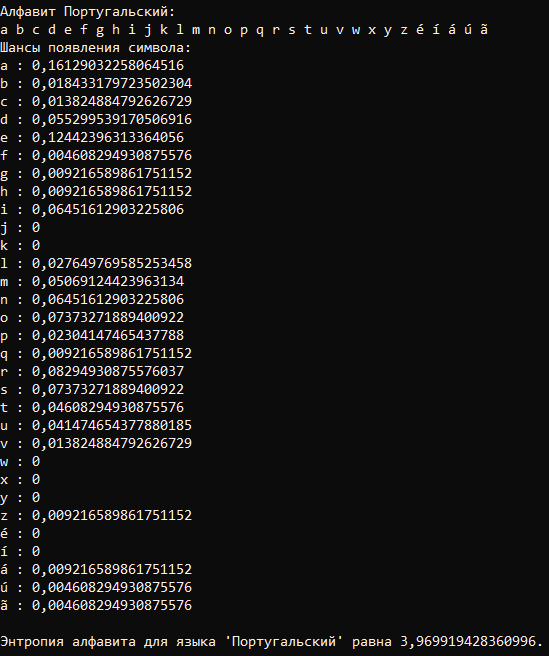


Рисунок 2 – вывод количества вхождений и энтропии португальского алфавита

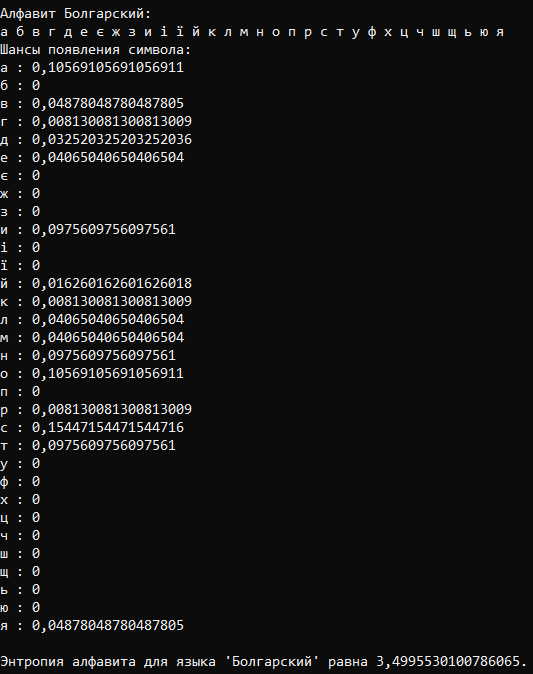


Рисунок 3 – вывод количества вхождений и энтропии болгарского алфавита



Рисунок 4­ – вывод суммы шансов для обоих языков

Приложение также генерирует таблицу в MSExcel на основании частот появления символов, рассчитанных на предыдущем шаге.

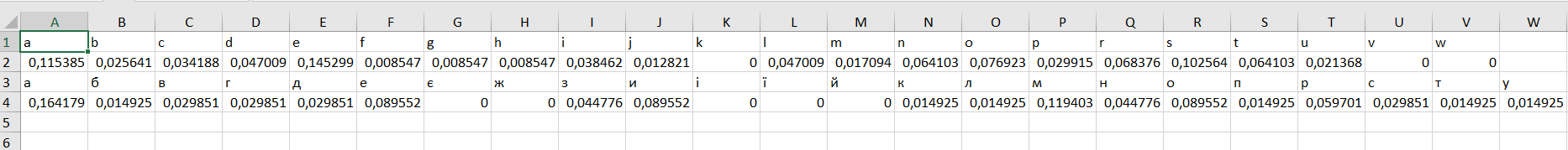


Рисунок 5­ – таблица частот появления символов

Затем приложение рассчитывает эффективную энтропию португальского и болгарского алфавита, рассчитанную по формуле , где , для различных вероятностей ошибочной передачи единичного бита (0.1, 0.5, 1.0).

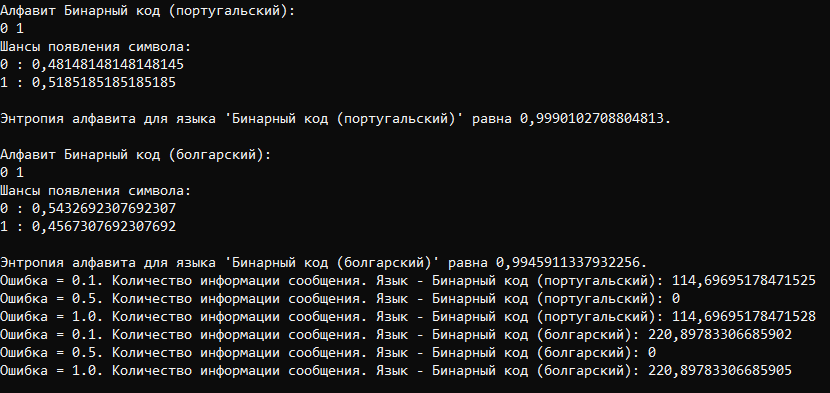


Рисунок 6­ – таблица частот появления символов

**Вывод**

В ходе лабораторной работы были закреплены теоретические знания по основам теории информации.

Также было разработано приложение для расчета энтропии по Шеннону и количества информации алфавитов (португальского, болгарского и бинарного).