УО «Белорусский государственный технологический университет»

Факультет информационных технологий

Кафедра программной инженерии

**ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ № 2**

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

по дисциплине

«Криптографические методы защиты информации»

Руководитель: Блинова Е.А.

Выполнил: студент 3 курса

5 группы Храмых В.О.

Минск 2020

**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.

**Краткие теоретические сведения**

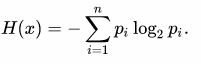
Алфавит источника сообщения – это конечная или бесконечная совокупность символов и знаков, формирующих сообщение.

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.

Мощность алфавита – это количество символов, составляющих алфавит. Информационной характеристикой алфавита (источника сообщений на основе этого алфавита) является энтропия.

Энтропия алфавита – это информационная характеристика алфавита. Энтропия характеризует количество информации, приходящееся в среднем на один символ алфавита. Зависит только от распределения вероятностей, а не от алфавита.

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



где pi – вероятность i-ого события.

С физической точки зрения энтропия алфавита 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:

*,*

где N – мощность алфавита.

Энтропия по Хартли рассчитывается с учетом того, что алфавит состоит из символов с равными вероятностями появления. Для расчета энтропии по Шеннону необходимо вычислить отдельно вероятность появления каждого символа алфавита.

**Ход работы**

Создать приложение для расчета и анализа параметров и информативных характеристик дискретных ИС, с помощью которого:

а) рассчитать энтропию указанного преподавателем алфавитов: один – на латинице, другой – на кириллице (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);

б) для входных документов, представленных в бинарных кодах, определить энтропию бинарного алфавита;

в) 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 алфавита – а) и в кодах ASCII – б); объяснить полученный результат;

г) выполнить задание п. в) при условии, что вероятность ошибочной передачи единичного бита сообщения составляет: 0.1; 0.5; 1.0.

Для выполнения задания необходимо задать символы 4 разных алфавитов, которые приведены на рисунке 1:

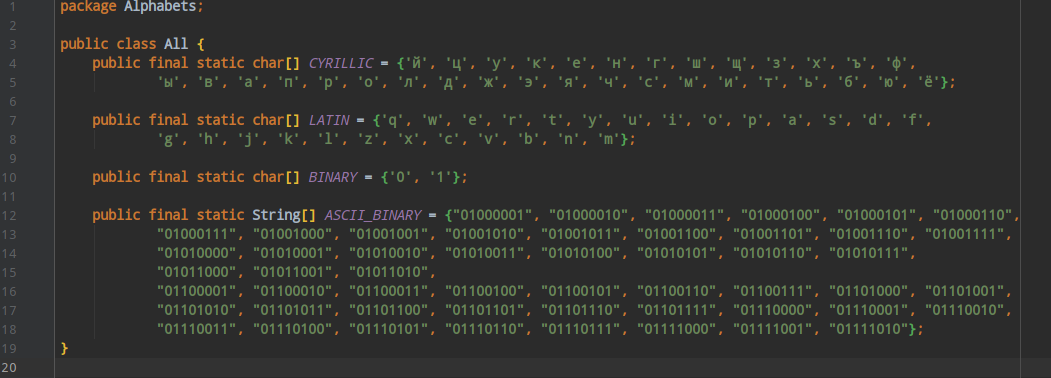


Рисунок 1 - Алфавиты

Для расчета энтропии алфавитов и расчета вероятностей были использованы стандартные формулы, приведенные в теоретических сведениях. В качестве входных документов для проведения точных расчетов были приняты текстовые документы для каждого из 4 алфавитов, также в приложении Excel были построены гистограммы, отражающие частоты появления символов.

На рисунке 2 представлен программный код, демонстрирующий алгоритм расчета энтропии и вероятностей алфавита:



Рисунок 2 – Алгоритм расчёта

Для расчёта энтропии текста с ошибкой используются следующие 2 метода.

Результат вычисления выдаёт метод getCountEffectiveInfo:

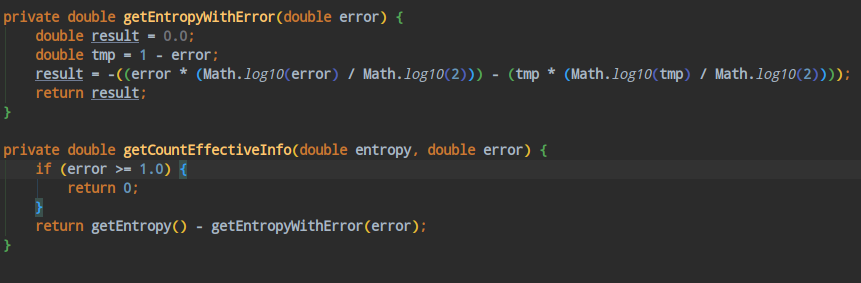


Рисунок 3 – Методы расчёта энтропии с ошибкой

Гистограммы (по оси x – вероятность, по y – элемент алфавита):

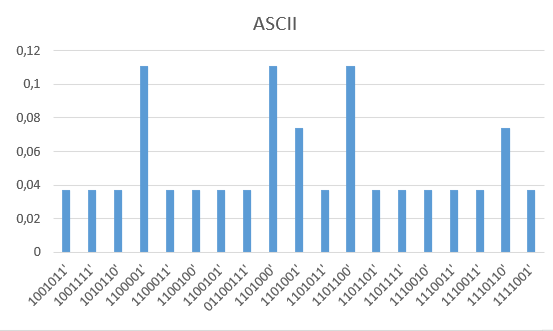
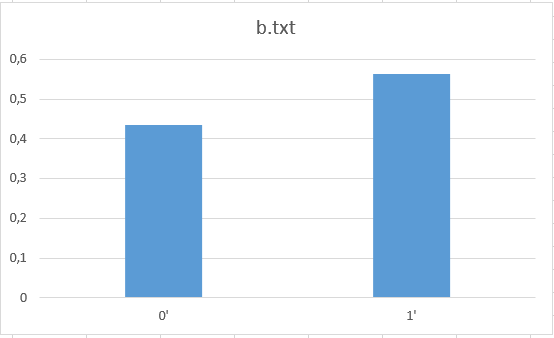
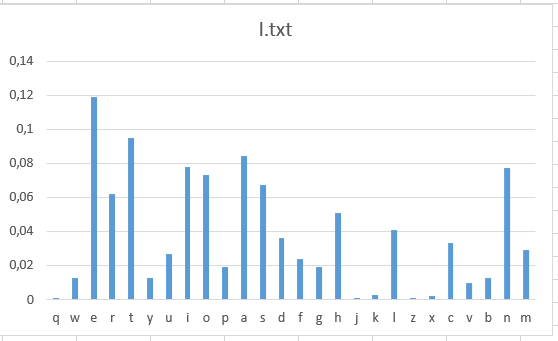
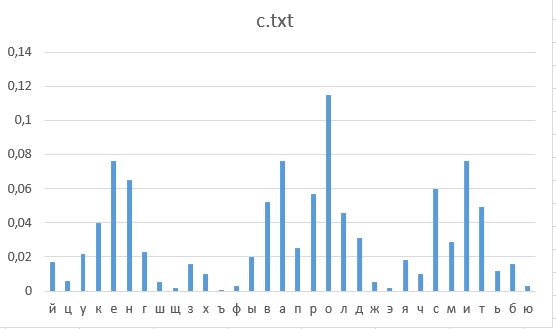
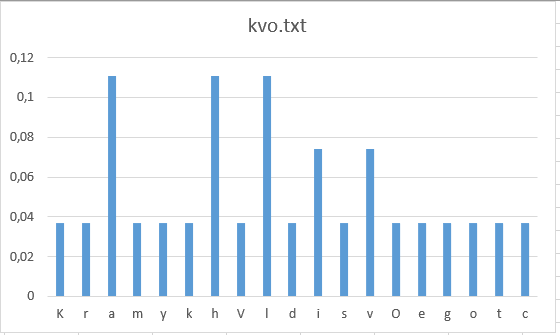
kvo.txt – ФИО на латинице;

c.txt – большой текст на КИРИЛЛИЦЕ;

l.txt – большой текст на ЛАТИНИЦЕ;

b.txt – большой текст в БИНРНОМ виде;

kvoASCII.txt (ASCII) – большой текст в бинарных кодах ASCII;



Результат выполнения приложения представлен в файле out.txt (он большой, поэтому счёл нецелесообразным его вставлять в этот документ)