Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Лабораторная работа №1

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

Студент: Подобед В.Г.

ФИТ 3 курс 6 группа

Преподаватель: Нистюк О.А.

Минск 2024

**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.

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.

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

**Задание 1:**

а) рассчитать энтропию указанных 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);

Для расчета энтропии голландского(латинице) алфавита и украинского алфавита(кириллица) с помощью разработанного нами приложения текст из входного файла text\_golland.txt и text\_ukraina.txt записывается в строку, которая затем передается в качестве параметра при вызове функции, рассчитывающей энтропию алфавита. Описание данной функции представлено на рисунке 1.1.

Рассчитанное по формуле Шеннона: .

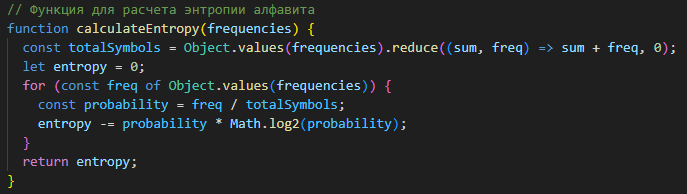


Рис. 1.1 – Функция для расчета энтропии

Результат выполнения данной функции выводится в консоль, содержимое которого представлено на рисунке 1.2 и на рисунке 1.3.

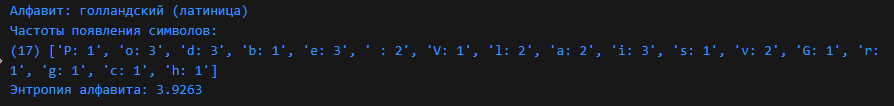


Рис. 1.2 –Результат расчета энтропии голландского языка

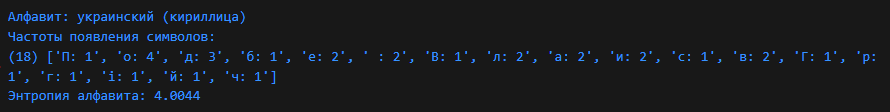


Рис. 1.3 – Результат расчета энтропии украинского языка

Также в MSExcel на основании частот появления символов строится гистограмма. Данная гистограммы голландского и украинского отображена на рисунке 1.4 и на рисунке 1.6. Функция для подсчёта чистоты представлена на рисунке 1.5.

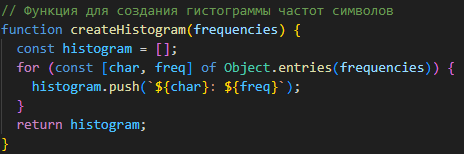


Рис. 1.4 – Гистограмма частот появления символов голландского языка



Рис. 1.4 – Гистограмма частот появления символов голландского языка



Рис. 1.6 – Гистограмма частот появления символов украинского языка

**Задание 2:**

Для входных документов был создан файл binary\_data.txt, представленных в бинарных кодах, определить энтропию бинарного алфавита. Результат выполнения выполнен на рисунке 2.1.

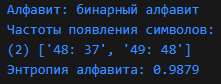


Рис. 2.1 – Результат расчета энтропии бинарного алфавита

**Задание 3:**

Используя значения энтропии алфавитов, полученных в пунктах (а) и (б), подсчитать количество информации в сообщении, состоящем из собственных фамилии, имени и отчества (на основе исходного алфавита – (а) и в кодах ASCII – (б)). Результат выполнения работы представлен на рисунке 3.1.

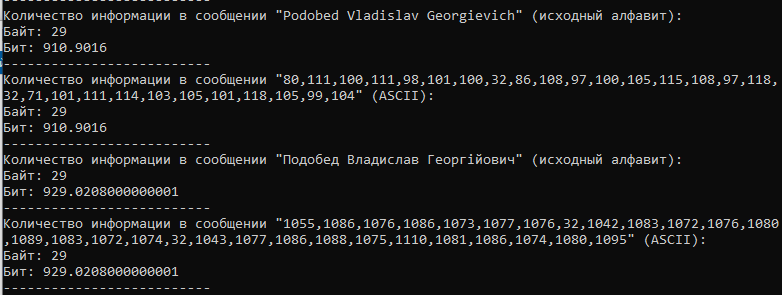


Рис. 3.1 – Результат

**Задание 4:**

Выполнить задание пункта (в) при условии, что вероятность ошибочной передачи единичного бита сообщения составляет: 0,1; 0,5; 1,0. Результат выполнения работы представлен на рисунке 4.1.

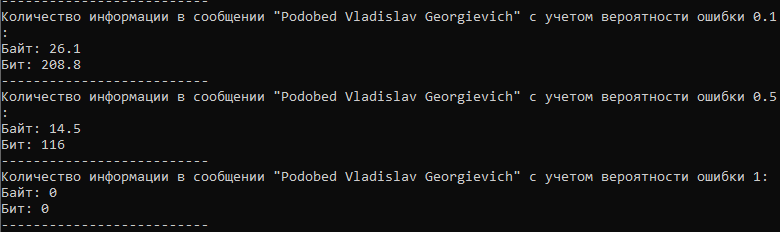


Рис. 4.1 – Вероятность ошибочной передачи

**Вывод**

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

Также было разработано приложение для расчета и анализа параметров и информативных характеристик дискретных ИС.