Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

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

Студент: Круглик Алексей

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)

# **Практические задания**

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

Задание 1:

рассчитать энтропию алфавитов:

* один – на латинице (испанский), другой – на кириллице (сербский);

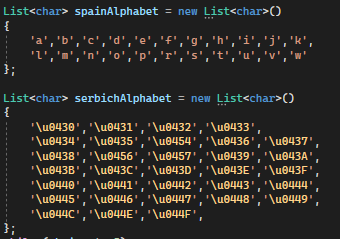


Рисунок 1 — Испанский и сербский алфавит

* в качестве входного принят произвольный электронный текстовый документ на основе соответствующего алфавита;

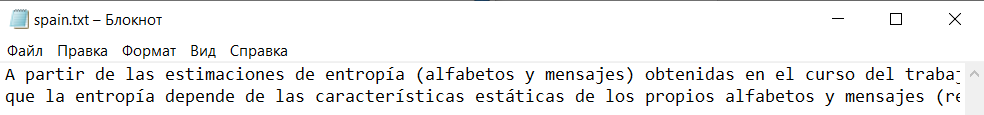


Рисунок 2 — Текстовый документ на испанском языке

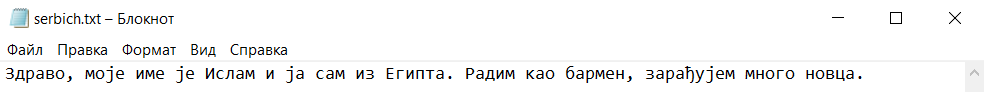


Рисунок 3 — Текстовый документ на сербском языке

* частоты появления символов алфавитов оформлены в виде гистограмм.

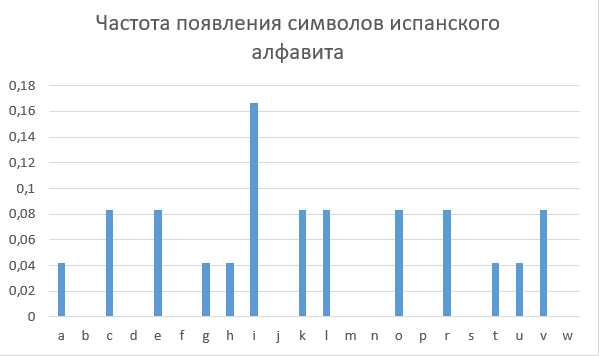


Рисунок 4 — Частота появления символов испанского алфавита



Рисунок 5 — Частота появления символов сербского алфавита





Рисунок 6 — Энтропия для испанского и сербского алфавитов

Задание 2:

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

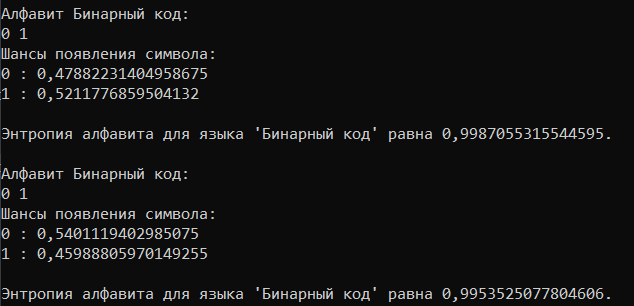


Рисунок 7 — Энтропия для бинарного алфавита

Задание 3:

подсчитать количество информации в сообщении, состоящем из собственных фамилии, имени и отчества; объяснить полученный результат;





Рисунок 8 — Количество информации для ФИО на основе испанского, сербского алфавитов и в кодах ASCII

Вероятность появления символов в кодах ASCII больше, чем в алфавитах испанского и сербского алфавитов, так как в кодах ASCII всего 2 символа, в испанском алфавите — 32 символа, в сербском — 33 символов. Соответственно, количество информации в кодах ASCII больше.

Задание 4:

выполнить задание предыдущего пункта при условии, что вероятность ошибочной передачи единичного бита сообщения составляет: 0,1; 0,5; 1,0.

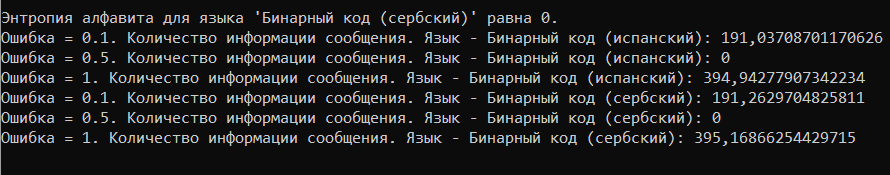


Рисунок 9 — Количество информации при условии, что вероятность ошибочной передачи единичного бита сообщения=0,1; 0,5; 1,0**Вывод**

В ходе выполнения лабораторной работы было создано приложение, позволяющее рассчитывать энтропию по Шеннону и количество информации алфавитов (испанского, сербского и бинарного). Если алфавит содержит больше двух символов, то нельзя использовать формулу нахождения эффективной энтропии, так как только в двоичном алфавите мы можем менять символ на противоположный при вероятности появления ошибки.