Министерство образования Республики Беларусь

Учреждение образования «Белорусский государственный технологический университет»

Факультет информационных технологий

Кафедра информатики и веб-дизайна

**Отчет к лабораторной работе**:

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

Выполнила:

студентка 3 курса 9 группы

специальности ДЭИВИ

Змитревич Д. А.

Минск 2023

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

**Задачи**:

1. Закрепить теоретические знания по основам теории инфор- мации.

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.

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.

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

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.

Алфавит, А– это общее число знаков или символов (N), используемых генерации или передачи сообщений. Символы алфавита будем обозначать через {, 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 К. Шенноном и Р. Хартли.

Энтропию алфавита А{} по Шеннону рассчитывают по следующей формуле:

(1.1)

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.

1. Практическая часть

а) Энтропия для итальянского алфавита – 4.39, монгольского – 5.13.



Рисунок 1 – Строка на латинице

Рисунок 2 – Гистограмма частоты символов



Рисунок 3 – Строка на кириллице

Рисунок 4 – Гистограмма частоты символов

б) Энтропия бинарного алфавита – 1.

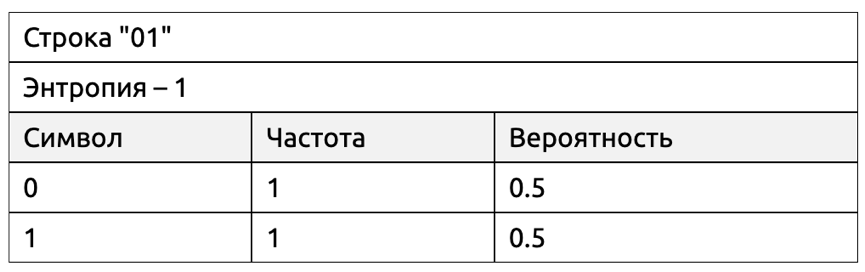


Рисунок 5 – Расчет энтропии бинарного алфавита

в) ФИО на латинице:Zmitrevich Diana Andreevna.

ФИО на кириллице:Змитревич Диана Андреевна.

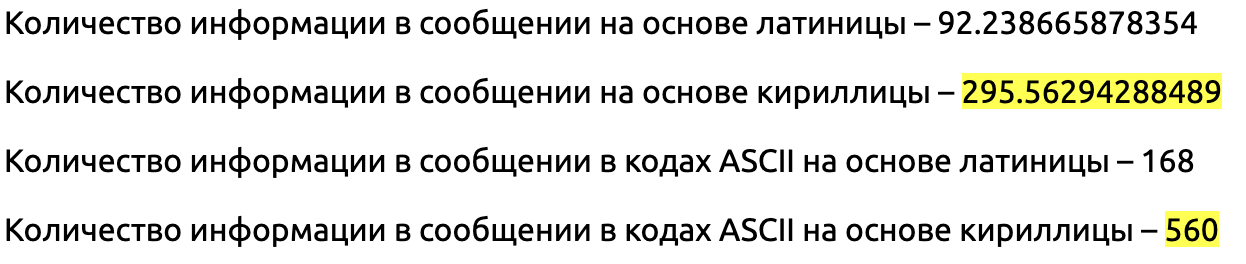


Рисунок 6 – Расчет количества информации (bit)

г) ФИО на латинице:Zmitrevich Diana Andreevna.

ФИО на кириллице:Змитревич Диана Андреевна.

С вероятностью ошибочной передачи 0.1:

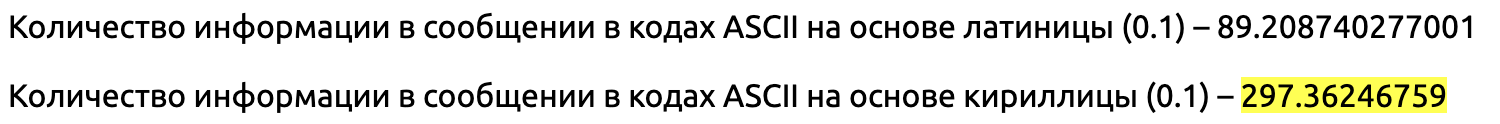


Рисунок 7 – Расчет количества информации (bit)

С вероятностью ошибочной передачи 0.5:

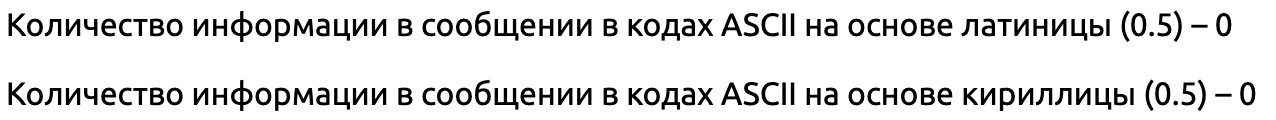


Рисунок 8 – Расчет количества информации (bit)

С вероятностью ошибочной передачи 1:

Количество информации в сообщении в кодах ASCII для английского алфавита: не число, т.к. логарифм от числа 0 по основанию 2 не существует (считается при вычислении условной энтропии).

Количество информации в сообщении в кодах ASCII для русского алфавита: не число, т.к. логарифм от числа 0 по основанию 2 не существует (считается при вычислении условной энтропии).

**Вывод**

В ходе лабораторной работы были приобретены практические навыки расчета и анализа параметров и информативных характеристик дискретных ИС. Также была разработана программа, вычисляющая энтропию латинского алфавита и кириллицы.

**Контрольные вопросы**

1. Что такое алфавит источника сообщения?

Это общее число знаков или символов (N), используемых для генерации или передачи сообщений.

1. Что такое мощность алфавита источника сообщения?

N – мощность алфавита.

1. Какова мощность алфавита белорусского языка?

N = 32.

1. Какова мощность алфавита русского языка?

N = 33.

1. Какова мощность алфавита «компьютерного» языка?
2. N = 2.
3. Что такое энтропия алфавита?

Количество значимой информации, приходящейся на один символ алфавита.

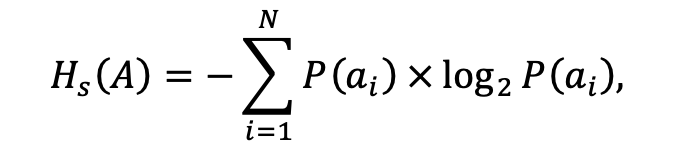
1. Что такое энтропия сообщения?

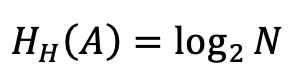
Количество значимой информации, приходящейся на один символ сообщения.

1. От чего зависит энтропия алфавита?

От мощности алфавита.

1. Записать формулу для вычисления энтропии.





1. Что нужно знать для вычисления энтропии алфавита?

Его мощность.

1. Как рассчитываются энтропия Шеннона и энтропия Хартли? В чем принципиальное различие между этими характеристиками? Дайте толкование физического смысла энтропии.

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.

1. Пояснить назначение знака «минус» в формулах (2.1) и (2.4).

Это объясняется тем, что вероятность по определению меньше или равна единице, а логарифм числа меньшего единицы – величина отрицательная.

1. Что такое избыточность алфавита и избыточность сообщений, сформированных в компьютерных системах? Принцип действия каких систем основан на существовании данной избыточности?

Избыточность сообщений – означает превышение количества информации, используемой для передачи или хранения сообщения, над его информационной энтропией.

Избыточность алфавита – показывает относительную недогруженность на символ алфавита.

1. Расположить в порядке возрастания энтропии известные Вам алфавиты.

Английский, русский.

1. Вычислить энтропию алфавита белорусского (русского) языка.

Энтропия русского языка – 5,04.

1. Вычислить энтропию Шеннона бинарного алфавита, если вероятность появления в произвольном документе на основе этого алфавита одного из символов составляет 0,25, другого – 0,75; либо 0 и 1,0; либо 0,5 и 0,5.

* 3.0195500086539,
* NAN
* 0

1. Чему равна энтропия алфавита по Хартли, если мощность этого алфавита равна: а) 1 символ; б) 2 символа; в) 8 символов?

* 0
* 1
* 3

Энтропия в криптографии – это мера случайности или неопределенности в пароле, ключе или других секретных данных. Чем выше энтропия, тем больше возможных комбинаций пароля или ключа, и тем сложнее его угадать или взломать.

В более простых словах, энтропия – это степень того, насколько ваш пароль или ключ являются сильными и защищенными от злоумышленников. Если вы используете короткий или очень простой пароль, то его энтропия будет низкой и он может быть легко угадан или взломан. Но если вы используете длинный и случайный пароль, то его энтропия будет высокой, и он будет надежным для защиты ваших секретных данных.