Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

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

Студент: Дрозд А. И.

ФИТ 3 курс 2 группа

Преподаватель:

Нистюк Ольга Александровна

Минск 2024

Лабораторная работа № 2

ЭЛЕМЕНТЫ ТЕОРИИ ИНФОРМАЦИИ. ПАРАМЕТРЫ И ХАРАКТЕРИСТИКИ ДИСКРЕТНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

**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

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.

**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.

Двоичный канал передачи информации строится на основе двоичного алфавита: А = {0, 1}. При этом канал, в котором вероятности искажения переданного 0 (принята соответственно 1; этому событию соответствует условная вероятность р(1|0)) и переданной 1 (принят соответственно 0; этому событию соответствует условная вероятность р(0|1)) равны, как и равны вероятности передачи 0 (р(0)) и 1 (р(1)), называют двоичным симметричным каналом (ДСК).

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

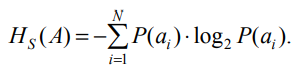
****

Рисунок 1.1 – Формула Энтропии алфавита по Шеннону

**С физической точки зрения энтропия алфавита показывает, какое количество информации приходится в среднем на один символ алфавита.**

Частным случаем энтропии Шеннона считается энтропия Хартли. Дополнительным условием при этом является то, что все вероятности одинаковы и постоянны для всех символов алфавита. С учетом этого формулу (2.1) можно преобразовать к виду:



Рисунок 1.2 – Частный случай формулы Энтропии алфавита

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

****

Рисунок 1.3 – Формула определяющая количество информации

Здесь Н(А) – энтропия алфавита с соответствующим распределением вероятностей р(аi).

Если принять, что р(аi = 1) = р(1) и р(аi = 0) = р(0), используя выражение (2.1), вычислим энтропию бинарного алфавита:



Рисунок 1.4 – Формула Энтропии бинарного алфавита

К примеру, если сообщение Хk состоит только из единиц (Хk = 11…1) и имеет длину k, то вероятность того, что произвольный символ равен единице, составляет единицу (р(аi = 1) = 1), и другая вероятность р(аi = 0) = 0 для i =1,N. Фактически здесь имеет место использование моноалфавита: алфавита, состоящего из одного символа.

Учитывая, что сумма р(1) + р(0) = 1, и выражая одну вероятность через другую (например, р(1) = 1 – р(0)), можно теоретически доказать информативность бинарного алфавита, решив дифференциальное уравнение [dH(A)/dp(1)] = 0 (вспомним из курса математики, как найти экстремум функции; можно для этого воспользоваться пособием [5]).

Если вероятность ошибки в ДСК отлична от 0 (р > 0), переданное сообщение может содержать ошибки: Хk ≠ Yk. Количество информации в таком сообщении при его передаче по ДСК будет определяться не энтропией двоичного алфавита (в соответствии с выражением (2.3)), а эффективной энтропией Hе(A) алфавита или пропускной способностью канала:



Рисунок 1.5 – Формула пропускной способности канала

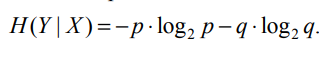


Рисунок 1.6 – Формула условной энтропии

**Практическое задание**

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

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);

Чтобы найти энтропию алфавитов необходимо использовать формулу, приведенную на рисунке 1.1. Для этого используется функция Entropy\_Shen, показанная на рисунке 1.6.

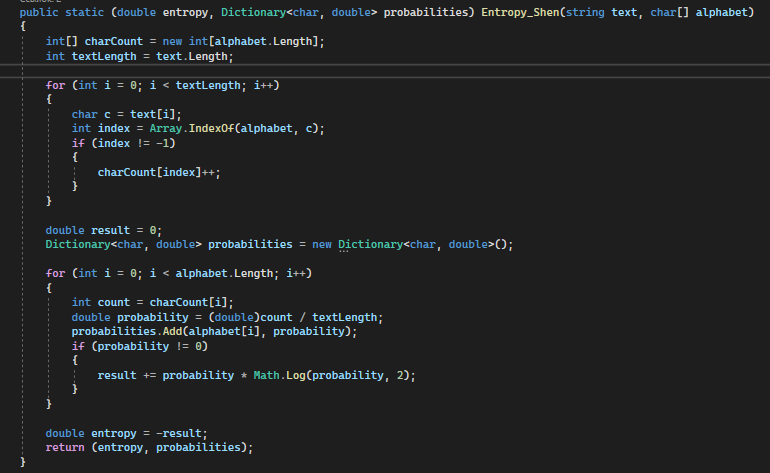


Рисунок 1.7 – Функция вычисления энтропии алфавита

Итогом выводится информация об общей энтропии заданного текста и все вероятности появления каждого символа указанного алфавита. Результат выполнения представлен на рисунке 1.8.

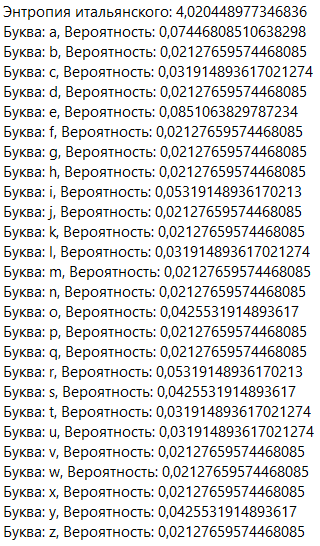


Рисунок 1.8 – Результат вывода на страницу

Также для наглядности необходимо было создать 2 гистограммы, которые можно увидеть на рисунке 1.8 и 1.9



Рисунок 1.9 – Диаграмма энтропии итальянского



Рисунок 1.10 – Диаграмма энтропии монгольского

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

Для выполнения данного задания необходимо перевести заданный текст в бинарный формат. Для этого будем использовать функцию Binary, которая преобразует символьную информацию в набор значений (1,0). Данная функция показана на рисунке 1.11.

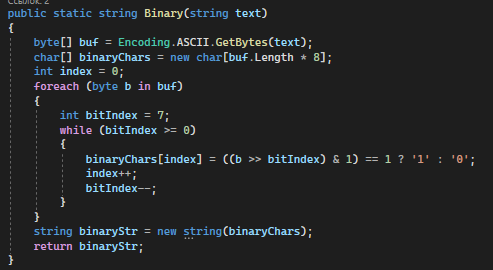


Рисунок 1.10 – Функция Binary

Результат нахождения после перевода текста в бинарную систему показан на рисунке 1.11.



Рисунок 1.11 – Результат нахождения

в) 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)); объяснить полученный результат;

Для выполнения данного задания необходимо было перевести ФИО в кодировку ASCII. Чтобы выполнить эту часть, был использован онлайн-конвертер текста в необходимую кодировку. В результате был получен набор последовательных числовых значений.

Далее была создана функция CountInfo, которая умножает кол-во символов на энтропию строки (показана на рисунке 1.12). Формула, которая использовалась в этом задании представлена на рисунке 1.3, тоже самое и для строки в кодировке ASCII, результат можно увидеть на рисунке 1.13.



Рисунок 1.12 – Функция CountInfo

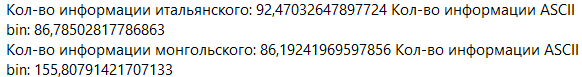


Рисунок 1.13 – Итоговый результат выполнения

Можно заметить, что кол-во информации в имени меньше, чем в кодировке ASCII, это можно объяснить тем, что кол-во символов в кодировке ASCII значительно больше, чем в обычном представлении, и значит, по формуле 1.3, можно сказать именно по какой причине кол-во информации больше в тексте по кодировке ASCII.

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

Вероятность ошибочной передачи единичного бита сообщения влияет на количество информации, которое можно получить из сообщения. Если вероятность ошибки равна 0, то каждый бит сообщения передается без ошибок, и мы получаем максимальное количество информации. Если вероятность ошибки равна 1, то каждый бит сообщения передается с ошибкой, и мы не получаем никакой информации.

Для вычисления значений вероятностей ошибочной передачи используется функция countInformationWithMistake, которая просто повторяет формулу 1.6, принимает параметр ошибочной передачи, подробнее можно увидеть на рисунке 1.14.

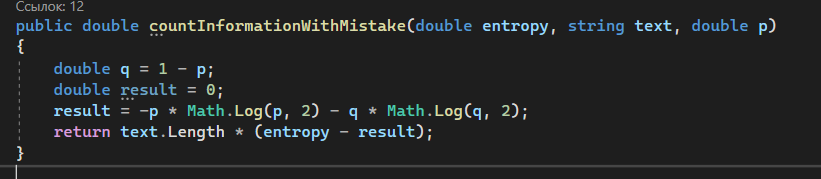


Рисунок 1.13 – Описание функции countInformationWithMistake

Далее нужно было последовательно применить это формулу к 3 значениям параметров (0.1, 0.5, 1), а также для 2 строк в кириллице и кодировке ASCII, результат можно увидеть на рисунке 1.14.

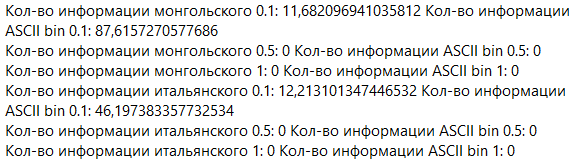


Рисунок 1.14 – Результат вероятности ошибочной передачи информации

В случае, когда вероятность ошибочной передачи единичного бита равна 0.5, количество информации равняется 0. Это обусловлено тем, что каждый второй бит будет неверный, получатель сообщения получит нечитаемое и непонятное сообщение, которое никак не соответствует первоначальному. В случае, когда вероятность ошибочной передачи единичного бита равна 1, количество информации также равняется 0. В данном случае каждый бит переданной информации будет неверным, соответственно мы получим полностью неверное сообщение.

**Вывод:** в ходе выполнения практических заданий было разработано приложение для анализа и расчета параметров дискретных информационных систем. Приложение позволяет эффективно рассчитывать энтропию алфавитов, как на итальянском, так и на монгольском языках, а также определять энтропию бинарного алфавита на основе входных данных в бинарном формате.

Проведено исследование, направленное на анализ количества информации в сообщении, которое включает в себя не только данные о фамилии, имени и отчестве, но и учитывает различные алфавиты, такие как исходный и алфавиты ASCII.

Дополнительно было изучено влияние вероятности ошибочной передачи отдельного бита сообщения на общее количество передаваемой информации. Полученные результаты предоставляют ценные инсайты для эффективной оценки и анализа информационных характеристик систем, что в свою очередь может сыграть ключевую роль в оптимизации и улучшении информационных процессов в широком спектре областей и приложений.