Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Дисциплина «Защита информации и надежность информационных систем»

**Отчёт по лабораторной работе №2**

Элементы теории информации. Параметры и характеристики дискретных информационных систем

Студент: Жук С.С.

ФИТ 3 курс 2 группа

Преподаватель: Савельева М.Г.

Минск 2025

**Содержание**

[1 Теоретические сведения 3](#_Toc193142185)

[2 Энтропия алфавита 5](#_Toc193142186)

[2.1 Кириллица 6](#_Toc193142187)

[2.2 Латиница 6](#_Toc193142188)

[2.3 Частота появления символов 7](#_Toc193142189)

[3 Количество информации 8](#_Toc193142190)

[4 Эффективная энтропия 9](#_Toc193142191)

[Вывод 11](#_Toc193142192)

# **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.

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.

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:

(1.1)

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:

(1.2)

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

(1.3)

Для ДСК: *p* + *q* = 1, где *q* – вероятность правильной передачи бита сообщения, а *p* – вероятность передачи бита с ошибкой. Если вероятность ошибки в ДСК отлична от 0 (*р* > 0), переданное сообщение может содержать ошибки: *Xk*≠ *Yk*. Количество информации в таком сообщении при его передаче по ДСК будет определяться не энтропией двоичного алфавита (в соответствии с выражением (1.3)), а эффективной энтропией алфавита или пропускной способностью канала:

(1.4)

где *H(Y|X)* – условная энтропия:

(1.5)

# **2 Энтропия алфавита**

Информационной характеристикой алфавита является энтропия. С физической точки зрения энтропия алфавита показывает, какое количество информации приходится в среднем на один символ алфавита. Энтропию алфавита *А*= {*ai*} по К. Шеннону рассчитывают по формуле 1.1.

Для вычисления энтропии в начале необходимо подсчитать количество появлений каждого символа в текстовом файле для каждого алфавита. Для этого используется функция, представленная в листинге 2.1. В ней мы инициализируем количество вхождений символов и после из текста файла перебираем каждый символ, добавляя его в такую структуру данных, как Map, хранящая данные в виде ключ-значение.

|  |
| --- |
| function calculateFrequencies(text, alphabet) {  const frequency = new *Map*();  let totalChars = 0;  for (const char of alphabet) {  frequency.set(char, 0);  }  for (const char of text) {  if (alphabet.includes(char)) {  frequency.set(char, frequency.get(char) + 1);  totalChars++;  }  }  return { frequency, totalChars }; } |

Листинг 2.1 – Функция подсчета количества появлений символов

Далее необходимо перейти от количества появлений символов к вероятности их появления. Для этого необходимо количество появлений каждого символа разделить на общее количество символов для каждого алфавита. Результат достигается при помощи функции, представленной в листинге 2.2.

|  |
| --- |
| function calculateProbabilities(frequency, totalChars) {  const probabilities = new *Map*();  for (const [char, count] of frequency.entries()) {  probabilities.set(char, count / totalChars);  }  return probabilities; } |

Листинг 2.2 – Функция вычисления вероятности

Далее эту вероятность *P*(*ai*)можно подставлять в формулу 1.1. Функция, вычисляющая энтропию Шеннона, представлена в листинге 2.3.

|  |
| --- |
| function calculateEntropy(probabilities) {  let entropy = 0;  for (const p of probabilities.values()) {  if (p > 0) {  entropy -= p \* *Math*.log2(p);  }  }  return entropy; } |

Листинг 2.3 – Функция вычисления энтропии Шеннона

## **2.1 Кириллица**

Для исследования кириллического алфавита был выбран текст на болгарском языке. Электронный текстовый документ на основе алфавита содержится в файле bulgarian.txt, представленном на рисунке 2.4.

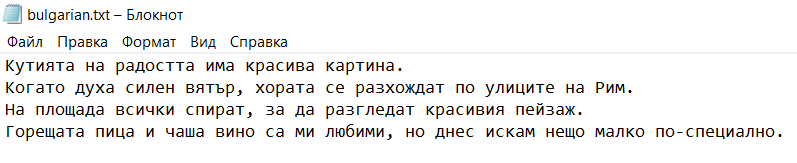


Рисунок 2.4 – Исходный текст болгарского алфавита

## **2.2 Латиница**

Для исследования алфавита на латинице был выбран итальянский алфавит. Электронный документ, в котором содержится текст на исходном итальянском алфавите для подсчета энтропии алфавита, содержится в файле italian.txt, представленном на рисунке 2.5.

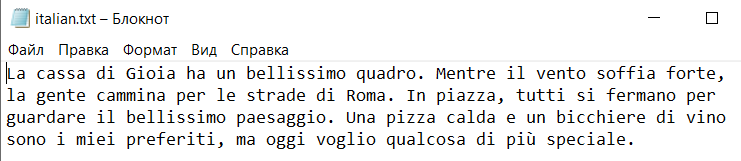


Рисунок 2.5 – Исходный текст итальянского алфавита

## **2.3 Частота появления символов**

Графики, которые отражают частоты появления символов в тексте итальянского и болгарского алфавита, представлены на рисунках 2.6 и 2.7 соответственно.

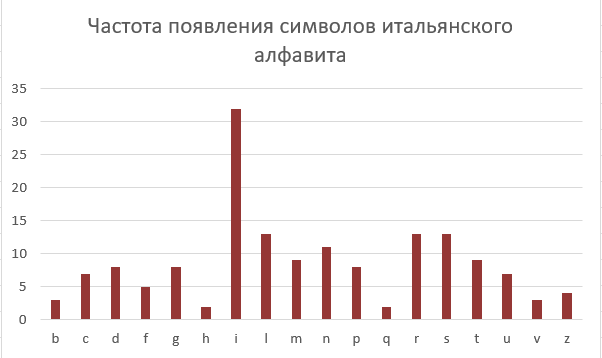


Рисунок 2.6 – Гистограмма частоты появления символов в тексте итальянского алфавита



Рисунок 2.7 – Гистограмма частоты появления символов в тексте болгарского алфавита

На основе этих данных можно получить вероятности появления каждого символа *P*(*ai*)*.* При подстановке их в формулу 1.1 получаем следующие значения энтропии для итальянского, болгарского и бинарного алфавитов. Энтропия итальянского и болгарского алфавитов равны соответственно 3,831 бит и 4,204 бит. Энтропия бинарного алфавита, которая всегда примерно равна 1 биту, в данном случае равняется 0,967 бит.

# **3 Количество информации**

Количество информации рассчитывается по формуле 1.3.

То есть для того, чтобы вычислить количество информации, необходимо умножить полученную энтропию алфавита на количество символов в сообщении. Расчет количества информации сообщения с помощью программы продемонстрирован в листинге 3.1.

|  |
| --- |
| function calculateInformation(entropy, messageLength) {  return entropy \* messageLength; } |

Листинг 3.1 – Функция расчета количества информации

Количество информации в сообщении итальянского, болгарского и бинарного алфавитов, состоящем из собственных фамилии, имени и отчества, равняется 80,443, 101,144, 212,636 битам соответственно.

# **4 Эффективная энтропия**

Если переданное сообщение может содержать ошибки, количество информации определятся для бинарного алфавита по формуле эффективной энтропии 1.4, а для небинарного по формуле 4.1:

(4.1)

В данной формуле *H(Y* | *X)* – условная энтропия, вычисляемая по формуле 1.5. Здесь *p –* вероятность передачи ошибочного символа.

Программный расчет эффективной энтропии представлен в листинге 4.1.

|  |
| --- |
| function calculateConditionalEntropy(p) {  const q = 1 - p;  return - p \* *Math*.log2(p) - q \* *Math*.log2(q); }  function calculateEffectiveEntropy(entropy, p, isBinary) {  if (p === 0.5) return 0;  if (isBinary && p === 1) return entropy;  else if (!isBinary && p === 1) return 0;  return entropy - calculateConditionalEntropy(p); } |

Листинг 4.1 – Функции расчета условной и эффективной энтропии

В результате получим количество информации для бинарного сообщения с различными вероятностями ошибок. Результат показан на рисунке 4.1.

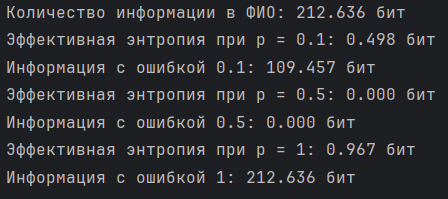


Рисунок 4.1 – Вывод терминала для бинарного алфавита

Рассчитаем то же самое, только для итальянского алфавита. Вывод продемонстрирован на рисунке 4.2.

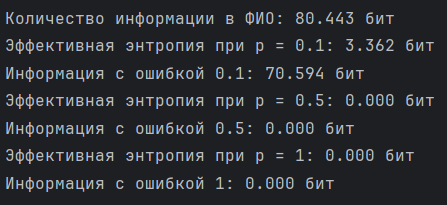


Рисунок 4.2 – Вывод терминала для итальянского алфавита

Как видно из данных рисунков, количество информации в исходном сообщении, состоящем из различных алфавитов, будет разное, и оно зависит от вероятности ошибки. В бинарном алфавите эффективная энтропия и количество информации будут равны длине строки тогда, когда вероятность ошибки равняется 0 или 1 соответственно. В небинарных алфавитах при вероятности ошибки, равной 1, эффективная энтропия и количество информации будут равняться 0.

# **Вывод**

В результате выполнения лабораторной работы были получены и закреплены практические навыки в области расчёта и анализа параметров и информативных характеристик дискретных информационных систем. Также были углублены теоретические знания, касающиеся основ теории информации.

Кроме того, было разработано специализированное приложение, которое позволяет эффективно осуществлять расчёты энтропии Шеннона и количества информации алфавитов (болгарского, итальянского и бинарного) и проводить анализ параметров и информативных характеристик дискретных информационных систем. Количество информации в сообщении зависит от энтропии алфавита: чем выше энтропия, тем больше информации несет каждый символ.

Также сделан вывод, что при вероятности ошибки, равной 1, для бинарного алфавита количество информации будет таким же, как при вероятности ошибки, равной 0, в виду того, что бинарный алфавит имеет свойство инвертированности, так как все биты исходного сообщения заменяются на обратные, поэтому из такого сообщения можно получить исходную информацию. Если алфавит содержит больше двух символов, то нельзя использовать формулу нахождения эффективной энтропии, так как мощность алфавита больше 2 и каждый символ заменяется на некоторые неизвестный символ из алфавита, поэтому в данном случае количество информации будет равняться 0.