Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Лабораторная работа №2

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

Студент: Сосновец М.И.

ФИТ 3 курс 4 группа

Преподаватель: Нистюк О.А.

Минск 2025

**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.

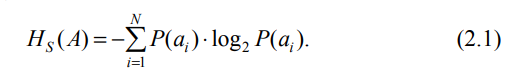
Алфавит, А – это общее число знаков или символов (N), используемых для генерации или передачи сообщений. Символы алфавита будем обозначать через {аi}, где 1 ≤ i ≤ N; N – мощность алфавита.

Минимальное число элементов алфавита Nmin = 2, А = {0, 1} – двоичный код. Один дискретный знак представляет собой элементарное сообщение, последовательность знаков – сообщение.

**Практическое задание**

**Задание 1**: 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 каждого языка (польского и белорусского) загружаются соответствующие текстовые файлы (polish.txt и belarus.txt), содержащие текст на этих языках. Затем для каждого текста рассчитываются частоты символов, принадлежащих соответствующему алфавиту (польскому или белорусскому). Каждый символ из текста проверяется, принадлежит ли он алфавиту, и если принадлежит, то увеличивается его счетчик в словаре. Энтропия текста рассчитывается для польского и белорусского языков. Этому процессу предшествует вычисление вероятностей появления каждого символа в тексте. Энтропия алфавита по К. Шеннону:



Реализация функции, которая рассчитывает энтропию текста представлена в листинге 1.1.

|  |
| --- |
| function entropyy(frequencies, totalChars) {      let entropy = 0;      for (const char in frequencies) {          const probability = frequencies[char] / totalChars;          entropy -= probability \* Math.log2(probability);      }      return entropy;  } |

Листинг 1.1 – Функция для расчета энтропии

Результат выполнения данной функции выводится в консоль, содержимое которого представлено на рисунке 1.1 и на рисунке 1.2.



Рисунок 1.1 – Энтропия польского языка



Рисунок 1.2 – Энтропия белорусского языка

В MS Excel на основе частоты появления символов строится гистограмма. Гистограммы для польского и белорусского языков представлены на рисунках 1.3 и 1.4 соответственно. Функция для подсчета частот приведена в листинге 1.2.



Рисунок 1.3 – Гистограмма текста латиницы



Рисунок 1.3 – Гистограмма текста кириллицы

Данная гистограмма показывает, что наиболее используемый символ «a».

|  |
| --- |
| function calculateFrequencies(text, alphabet) {      const frequencies = {};      const filteredText = text.split("").filter((char) => alphabet.includes(char));      filteredText.forEach((char) => {          frequencies[char] = (frequencies[char] || 0) + 1;      });      return frequencies;  } |

Листинг 1.2 – Функция для подсчета частоты символов

**Задание 2**: для входных документов, представленных в бинарных кодах, определить энтропию бинарного алфавита. Результат выполнения изображен на рисунках 1.4, 1.5 соответственно. Реализация функции показана в листинге 1.3.



Рисунок 1.4 – Результат расчета энтропии бинарного алфавита (белорусский язык)



Рисунок 1.5 – Результат расчета энтропии бинарного алфавита (польский язык)

|  |
| --- |
| function analyzeBinaryFile(filePath) {      try {          const buffer = fs.readFileSync(filePath);          let binaryString = buffer.toString('hex');          binaryString = binaryString.split("").map(char => parseInt(char, 16).toString(2).padStart(4, '0')).join("");          const frequencies = calculateFrequencies(binaryString, "01");          const totalChars = Object.values(frequencies).reduce((sum, freq) => sum + freq, 0);          console.log("частоты символов:", frequencies);          const entropy = entropyy(frequencies, totalChars);          console.log("этропия бинарного алфавита:", entropy.toFixed(4));      } catch (error) {          console.error("ошибка чтения бинарного файла:", error.message);      }  } |

Листинг 1.3 – Функция для подсчета 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)); объяснить полученный результат. Результат выполнения изображен на рисунке 1.6.



Рисунок 1.6 – Количество информации

Как видно из данного рисунка, количество информации в сообщении, состоящем из ФИО в кодировке ASCII, равняется 242 битам.

**Задание 4**: выполнить задание пункта (в) при условии, что вероятность ошибочной передачи единичного бита сообщения составляет: 0,1; 0,5; 1,0. Результат выполнения представлен на рисунке 1.7.

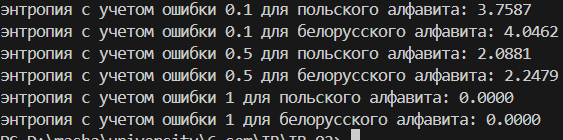


Рисунок 1.7 – Вероятность ошибочной передачи

**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.