**Лабораторная работа №2**

ЭЛЕМЕНТЫ ТЕОРИИ ИНФОРМАЦИИ. ПАРАМЕТРЫ И ХАРАКТЕРИСТИКИ ДИСКРЕТНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

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

**Теоретические сведения**

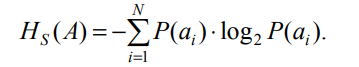
Передача информации (данных) осуществляется между двумя абонентами, называемыми источником сообщения (ИcС) и получателем сообщения (ПС). Третьим элементом информационной системы является канал (среда) передачи, связывающий ИсС и ПС. Отметим также, что и в системах с хранением информации всегда можно 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 какойлибо 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.

Рассмотрим основные характеристики и параметры двоичных систем. 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 знак представляет собой элементарное сообщение, последовательность знаков – сообщение. Набор элементов алфавита, создаваемых дискретным источником сообщений, заранее, априори (до опыта) известен получателю. Элементы теории информации. Параметры и характеристики дискретных систем 23. ИсС в каждый дискретный момент времени выдает один элемент алфавита. Этот элемент сообщения является одним из символов алфавита. Понятно, что ПС заранее не известно, какой это элемент. Если обозначить вероятность выбора каждого элемента алфавита p(аi), то 

Вероятности p(аi) могут быть получены в результате анализа частотных свойств символов алфавита, если на входе такого анализатора принять документ на основе соответствующего алфавита. Причем объем документа должен быть таким, чтобы от частости (частоты) появления каждого символа в анализируемом документе можно было перейти к вероятности соответствующего события.

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



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:



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



**Практическое задание**

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);

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

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;

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

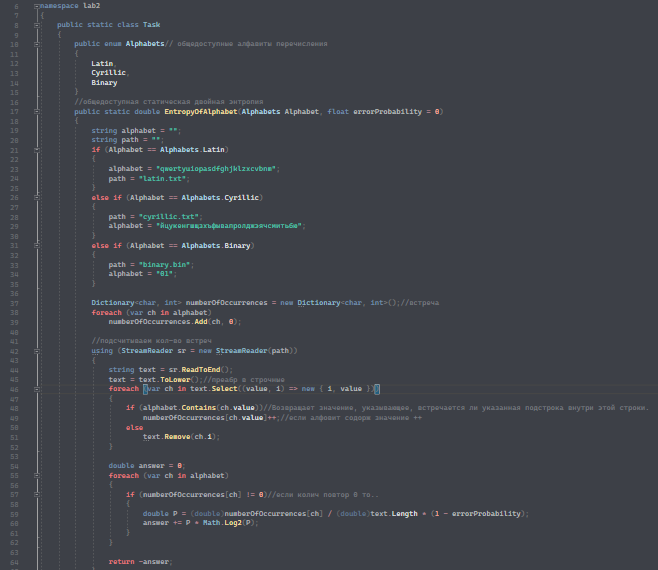


Рис. *1* – Код файла Task.cs

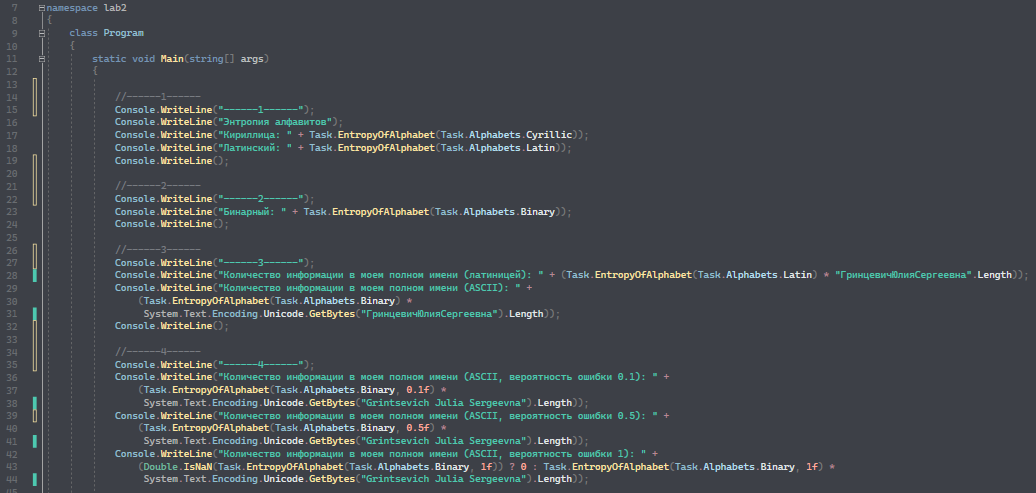


Рис. *2* – Код файла Program.cs

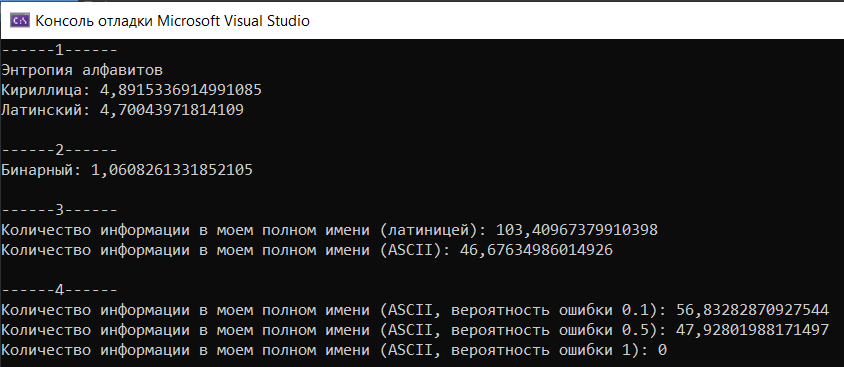


Рис. *3* – Результат выполнения программы

**Ответы на вопросы:**

1. **Что такое алфавит источника сообщения?**

**Алфавит** – конечная совокупность символов (знаков), с помощью которых можно представить любое сообщение в ИС.

1. **Что такое мощность алфавита источника сообщения?**

**Мощность алфавита** – количество символов, составляющих алфавит

1. **Какова мощность алфавита белорусского языка?**

32

1. **Какова мощность алфавита русского языка?**

33

1. **Какова мощность алфавита «компьютерного» языка?**

256

1. **Что такое энтропия алфавита?**

Информационной характеристикой алфавита (источника сообщений на основе этого алфавита) является *энтропия*. С физической точки зрения энтропия показывает, какое количество информации (бит) приходится в среднем на один символ алфавита.

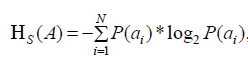
1. **От чего зависит энтропия алфавита?**

От частоты встречаемости символов и мощности алфавита.

1. **Что такое энтропия сообщения?**

Энтропия – информационная характеристика алфавита (источника сообщений на основе этого алфавита), которая показывает, какое количество информации приходится в среднем на один символ алфавита (сообщения).

1. **Записать формулу для вычисления энтропии (рисунок 1).**

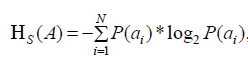


1. **Что нужно знать для вычисления энтропии алфавита?**

Частоту встречаемости каждого символа алфавита и мощность.

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

Энтропию алфавита А={*ai*} по К.Шеннону рассчитывают по следующей формуле, представленной на рисунке 2.



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 (2.1) можно преобразовать к виду, представленному на рисунке 3.



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

Поскольку мы находим логарифмы от чисел меньших 1, то логарифм всегда будет отрицательным, поэтому мы добавляем минус перед логарифмом

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

Избыточностью алфавита называется уменьшение информационной нагрузки на один символ вследствие неравновероятности и взаимозависимости появления его символов.

Информационная избыточность характеризует относительную нагруженность алфавита.

1. **Расположите в порядке возрастания энтропии известные вам** **алфавиты.**

3,2 – белорусского

3,895 – французского

4,25 – молдавского

1. **Вычислить энтропию алфавита белорусского (русского) языка.**

3.2 бит

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.**
2. 0,811 бит
3. 0 бит
4. 1 бит
5. **Чему равна энтропия алфавита по Хартли, если мощность этого алфавита равна: а) 1 символ, б) 2 символа, в) 8 символов?**
6. 0 бит
7. 1 бит
8. 3 бит

**Вывод:** приобрела практические навыки расчета и анализа параметров и информативных характеристик дискретных ИС.