Лабораторная работа № 1

по ТИПиС

студента группы ИТ-32

Манукова Давида Альбертовича

Выполнение: \_\_\_\_\_\_\_\_\_\_ Защита: \_\_\_\_\_\_\_\_\_\_

**Исследование избыточности источника информации**

Задание к работе

Задание А.

А.1. Получить (найти в Интернете) текстовые файлы, содержащие текст на одном из естественных языков или языке эсперанто, в соответствии со своим вариантом. Длина текста – не менее 10 тыс. символов.

А.2. Рассчитать оптимальную энтропию для заданного языка.

А.3. Используя программу «Исследование текста», выполнить расчет реальной энтропии языка на основании для полученного текстового файла.

А.4. Самостоятельно рассчитать избыточность сообщения заданного естественного языка при условии независимости символов алфавита.

А.5. Используя программу «Исследование текста», рассчитать избыточность сообщения заданного естественного языка при условии попарной зависимости символов алфавита.

А.6. Используя текстовый файл из п. А.1, сформировать усеченный алфавит языка, т.е. оставить в текстовом файле только символы из трети исходного алфавита. Пробелы и знаки препинания не учитывать, прописные и строчные символы не различать. Мощность усеченного алфавита должна быть не более 15.

А.7 – А.9. Выполнить пп. А.2 – А.4 для усеченного алфавита.

А.10. Построить таблицу частот символов первичного усеченного алфавита

А.11. Закодировать символы усеченного алфавита равномерным бинарным кодом.

Задание Б.

Б.1. Составить программу для экспериментального определения реальной энтропии сообщения при условии независимости символов. Выполнить расчет для полученного файла.

Б.2. Рассчитать избыточность сообщения заданного естественного языка с учетом попарной зависимости символов алфавита.

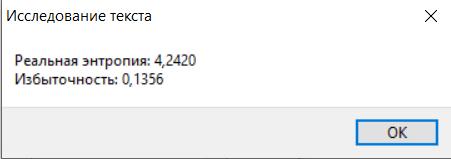
Б.3. Вместе с товарищами сравнить избыточность различных языков.

Б.4. Выполнить пункт Б.1 для усечённого алфавита.

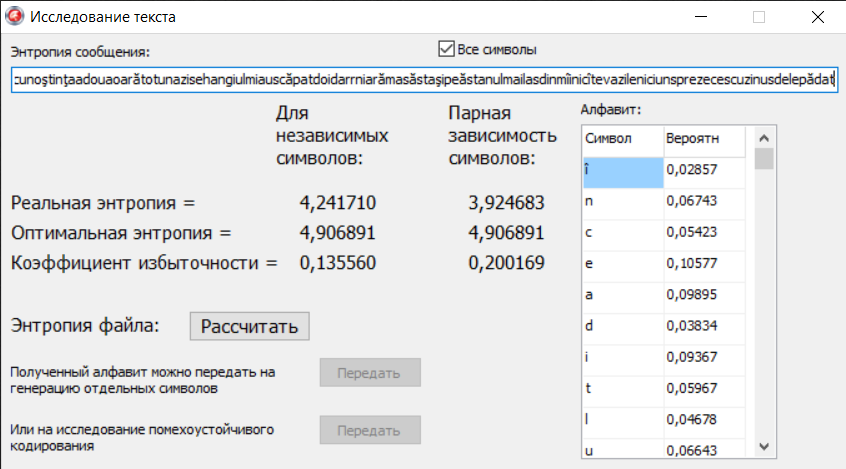
Ход работы

Задание А

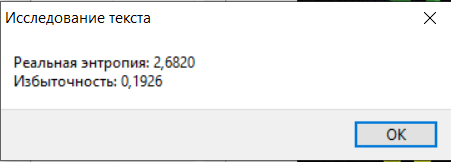
1. Нашел в Интернете текст на румынском языке.
2. Оптимальная энтропия: log228 = 4.807
3. Реальная энтропия: 4.242



1. Коэффициент избыточности: φ = 1 – 4,425/4,807 = 0,117
2. Используя программу «Исследование текста», рассчитал избыточность сообщения заданного естественного языка при условии попарной зависимости символов алфавита.



1. Укоротил алфавит текста до 10 символов.
2. Оптимальная энтропия = log210 = 3.322
3. Реальная энтропия



1. Коэффициент избыточности: φ = 1 – 2,682/3,322 = 0,192
2. Таблица частот символов для усеченного алфавита

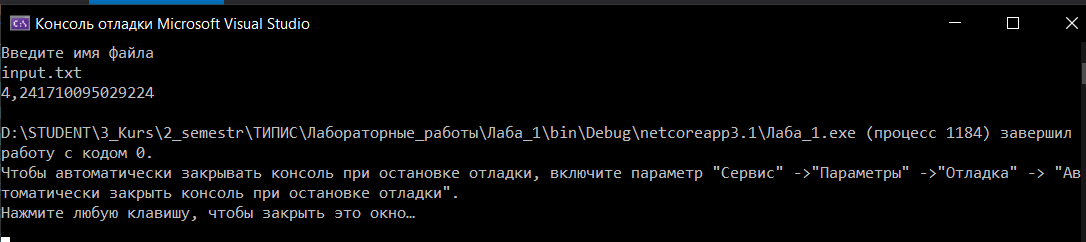
|  |  |
| --- | --- |
| e | 0.24415 |
| a | 0.22843 |
| d | 0.08849 |
| i | 0.21624 |
| g | 0.02974 |
| c | 0.12506 |
| f | 0.02401 |
| b | 0.02572 |
| h | 0.01304 |
| j | 0.00512 |

1. Закодировал равномерным бинарным кодом

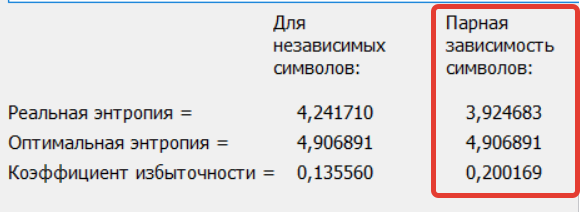
|  |  |
| --- | --- |
| e | 0000 |
| a | 0001 |
| d | 0010 |
| i | 0011 |
| g | 0100 |
| c | 0101 |
| f | 0110 |
| b | 0111 |
| h | 1000 |
| j | 1001 |

Задание Б

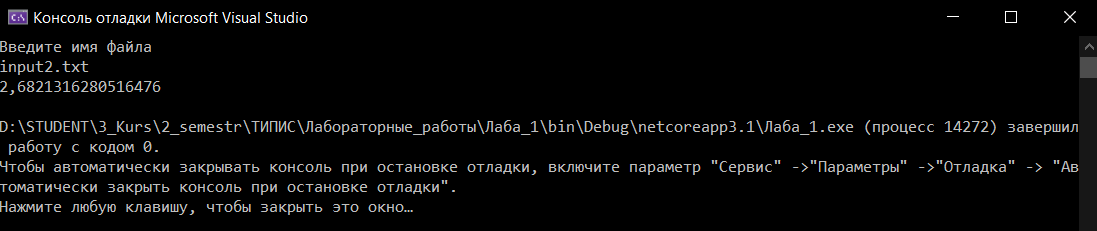
1. Написал программу для расчета энтропии сообщения при условии независимости символов.



1. Избыточность сообщения с учетом попарной зависимости символов.



1. Энтропия для усеченного алфавита



Вывод: В ходе в ходе выполнения лабораторной работы были приобретены навыки расчета характеристик источника информации, определения энтропии и избыточности естественных языков, первичного кодирования алфавита источника.