Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

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

Студент: Рубашек А. А.

ФИТ 3 курс 5 группа

Преподаватель:

Савельева М. Г.

1. **Цель работы**

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

а) рассчитать энтропию указанного преподавателем алфавитов: один – на латинице, другой – на кириллице ( 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 MSExcel);

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

в) используя значения энтропии алфавитов, полученных в пп. а) и б), подсчитать количество информации в сообщении, состоящем из собственных фамилии, имени о отчества (на основе исходного алфавита – а) и в кодах ASCII–б); объяснить полученный результат;

г) выполнить задание п. в) при условии, что вероятность ошибочной передачи единичного бита сообщения составляет: 0.1; 0.5; 1.0.

1. **Ход работы**

Для выполнения первого и второго задания было выбрано три алфавита: чешский, исландский и бинарный. Код этих алфавитов представлен на рисунке 2.1.

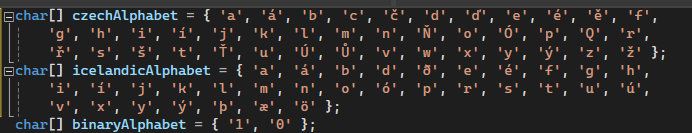


Рисунок 2.1 – Алфавиты

Для расчета энтропии алфавитов была разработана функция CalculateShannonEntropy(string text, char[] alphabet), которая рассчитывает энтропию по формуле Шеннона:

Код этой функции представлен на рисунке 2.2.

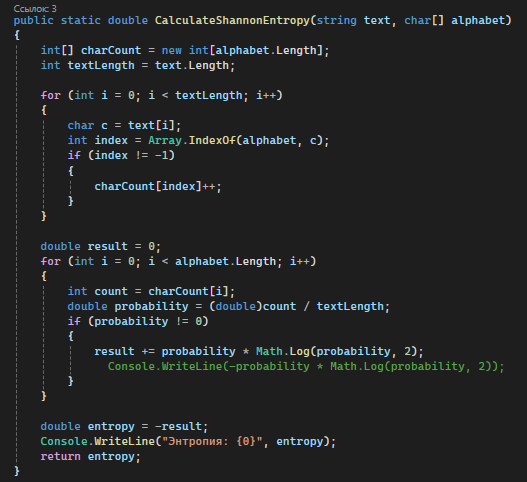


Рисунок 2.2 – Функция для расчёта энтропии по Шеннону

В результате получаем такие результаты расчёта энтропии чешского и исландского алфавитов для исходных текстов, которые представлены на рисунках 2.3 и 2.4



Рисунок 2.3 – Энтропия чешского алфавита



Рисунок 2.4 – Энтропия исландского алфавита

На основании частот появления символов, рассчитанных в предыдущем шаге, созданы гистограммы. Данные гистограммы отображены на рисунках 2.5 и 2.6.

Рисунок 2.5 – Гистограмма частот появления символов для чешского алфавита

Рисунок 2.6 – Гистограмма частот появления символов для исландского алфавита

Для расчёта энтропии для бинарного алфавита нужно было перевести текст в бинарный текст. Для этого была разработана функция Binary(string text), код которой представлен на рисунке 2.7.

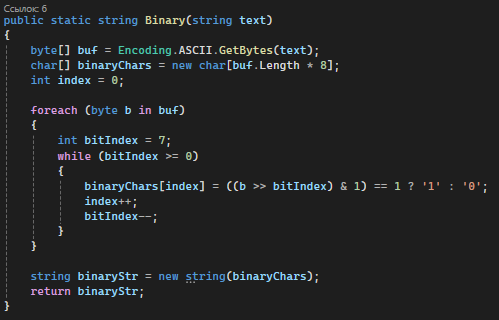


Рисунок 2.7 – Функция для перевода текста в бинарное представление

Следующим шагом нужно было рассчитать количество информации в сообщении, состоящем из собственных фамилии, имени о отчества. Для этого была разработана функция CountInformation(string text, double entropy), основанная на формуле *I*(*Xk*) *= H*(*A*)*⋅ k*. Её код представлен на рисунке 2.8.

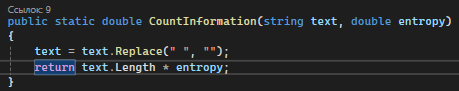


Рисунок 2.8 – Функция для расчёта количества информации

В результате получаем следующий вывод:

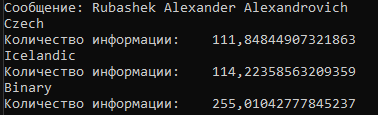


Рисунок 2.9 – Расчёт количества информации в сообщении

Следующим шагом нужно было подсчитать количество информации для бинарного алфавита при условии, что вероятность ошибочной передачи единичного бита сообщения составляет 0.1, 0.5, 1.0.

Поскольку вероятность ошибки отлична от нуля, то переданное сообщение может содержать ошибки. Количество информации в таком сообщении будет определяться не энтропией двоичного алфавита, а эффективной энтропией алфавита или пропускной способностью канала, которую можно рассчитать по следующей формуле:

,

где .

В результате получим следующий вывод:

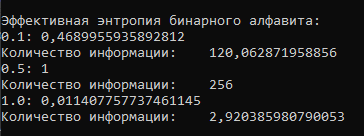


Рисунок 2.10 – Вывод терминала для бинарного алфавита

Затем рассчитаем то же самое, только для чешского алфавита. Получаем следующий вывод:

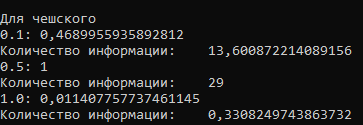


Рисунок 2.11 – Вывод терминала для чешского алфавита

В данном случае мы наблюдаем интересное явление: при использовании бинарного алфавита значение количества информации достигает максимума (равного 1) при вероятности ошибки равной 0.5. Это связано с особенностью бинарного алфавита, который имеет свойство инвертированности. Поэтому результаты для вероятностей ошибки 0.0 и 1.0 будут одинаковыми. Однако, когда вероятность ошибки стремится к 1 (например, 0.999), количество информации стремится к 0, что логично, поскольку в этом случае предсказать правильный бит становится практически невозможно.

**Вывод**

В ходе лабораторной работы были приобретены практические навыки расчёта и анализа параметров и информативных характеристик дискретных ИС, закреплены теоретические знания по основам теории информации.

Также было разработано приложение для расчета и анализа параметров и информативных характеристик дискретных ИС.