Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

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

Студент: Скалкович С.Л.

ФИТ 3 курс 5 группа

Преподаватель:

Нистюк Ольга Александровна

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 (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 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: таджикский, латышский и бинарный. Код этих алфавитов представлен на рисунке 2.1.

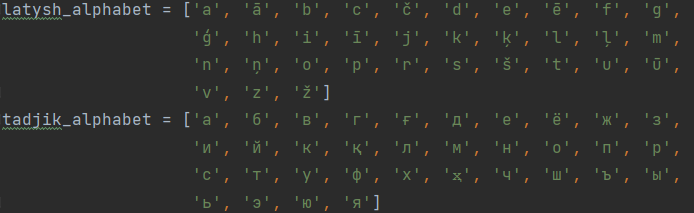


Рисунок 2.1 – Алфавиты

Для расчета энтропии алфавитов была разработана функция calculate\_shannon\_entropy(text, alphabet), которая рассчитывает энтропию по формуле Шеннона: . Код этой функции представлен на рисунке 2.2.

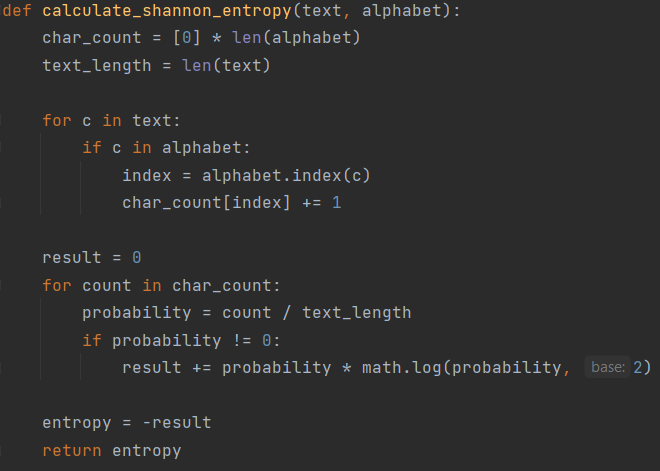


Рисунок 2.2 – Функция для расчёта энтропии по Шеннону

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



Рисунок 2.3 – Энтропия таджикского алфавита



Рисунок 2.4 – Энтропия латышского 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 таджикского алфавита

Рисунок 2.6 – Гистограмма частот появления символов для латышского алфавита

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

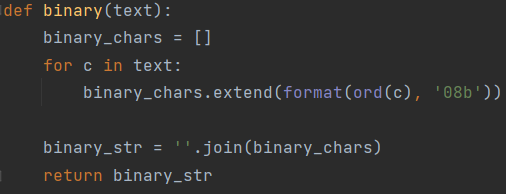
****

Рисунок 2.7 – Функция для перевода текста в бинарное представление

Результат использования функции calculate\_shannon\_entropy вместе с binary представлен на рисунке 2.8.



Рисунок 2.8 – Энтропия для бинарного алфавита

Следующим шагом нужно было рассчитать количество информации в сообщении, состоящем из собственных фамилии, имени о отчества. Для этого была разработана функция count\_information основанная на формуле *I*(*Xk*) *= H*(*A*)*⋅ k*. Её код представлен на рисунке 2.9.

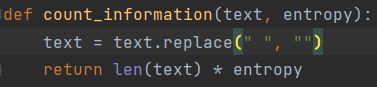


Рисунок 2.9 – Функция для расчёта количества информации

В результате получаем следующий вывод:

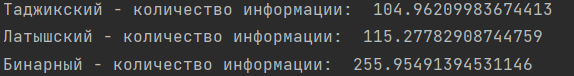


Рисунок 2.10 – Расчёт количества информации в сообщении

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:

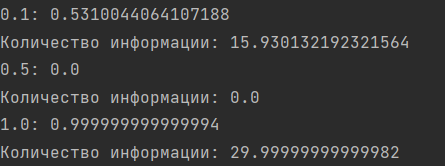


Рисунок 2.11 – Вывод терминала для бинарного алфавита

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

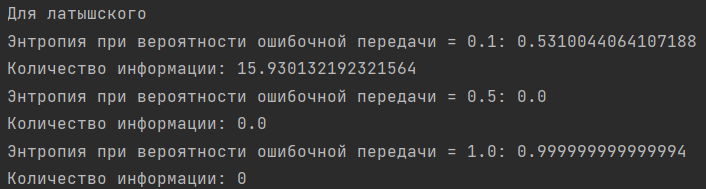


Рисунок 2.12 – Вывод терминала для латышского алфавита

В данном случае, когда мы используем бинарный алфавит, результат получается противоположным. При вероятности ошибки равной 1 мы получаем количество информации равное 1, а при вероятности ошибки равной 0 количество информации становится равным 0. Это объясняется свойством инвертированности бинарного алфавита, который обладает особенностью, что при вероятности ошибки 0 или 1 результат будет одинаковым. Пример: X = 0010010 – изначальное сообщение, при передаче его, при вероятности ошибки 100% мы получим 1101101, зная, что вероятность 100% мы сможем интерпретировать биты и получить изначальное сообщение.

**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 и проводить анализ параметров и информативных характеристик дискретных информационных систем.