МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Факультет информационных технологий

Кафедра информационных систем и технологий

Отчет по лабораторной работе №3

«ЭЛЕМЕНТЫ ТЕОРИИ ИНФОРМАЦИИ.ИНФОРМАТИВНОСТЬ ДАННЫХ  
В РАЗЛИЧНЫХ КОДИРОВКАХ»

по дисциплине

«Защита информации и надежность информационных систем»

Студент Шедько Е. А.

3 курс 3 группа ФИТ

Преподаватель: Берников Владислав Олегович

Минск 2021 г.

# Цель:

# Приобретение практических навыков трансформации данных и сопоставление энтропийных свойств используемых при этом алфавитов.

# Задачи:

# 1. Закрепить теоретические знания по взаимной конвертации данных, представленных в кодах ASCII и base64.

# 2. Разработать приложение для конвертации произвольного документа в формат base64 и обратно.

# 3. Исследовать энтропийные характеристики используемых в конвертерах алфавитов.

# 4. Изучить особенности практической реализации операции XOR над данными, представленными в разных форматах.

# 5. Результаты выполнения лабораторной работы оформить в виде описания разработанного приложения, методики выполнения экспериментов с использованием приложения и результатов эксперимента.

# Теоретическая часть:

Из энтропийных оценок мы выяснили, что энтропия зависит от статических характеристик самих алфавитов и сообщений. Энтропия максимальна при равномерном появлении букв на любом месте сообщения. Для характеристики источника сообщений с различным алфавитом введено понятие избыточности источника сообщений, или избыточности алфавита. Избыточностью алфавитаназывают уменьшение информационной нагрузки на один символ вследствие разной вероятности и взаимозависимости появления его символов в сообщениях. В наиболее общем виде избыточность алфавита *R* можно оценить отношением энтропии по Хартли и по Шеннону.

Кодировка base64 используется для кодирования бинарных данных.  
Кодирование base64 разработано для представления произвольных последовательностей октетов в форме, позволяющей использовать строчные и прописные буквы. Процесс кодирования представляет группу из 24 последовательных битов в форме строки из 4 символов. Обработка выполняется слева направо, а 24-битная исходная группа образуется конкатенацией трех 8-битных групп (байтов). Данные 24 бита после этого трактуются как 4 сцепленных группы по 6 битов, каждая из которых транслируется в один символ алфавита base64.  
Каждая 6-битная группа используется в качестве индекса массива из 64 печатных символов. Символы алфавита, соответствующие индексу, помещаются в выходную строку.

Если кодируются только один или два байта, в результате  
получаются только первые два или три символа строки, а выходная строка дополняется двумя или одним символами «=». Это предотвращает добавление дополнительных битов к восстановленным данным. Если на входе доступно менее 24 битов, входная группа дополняется (справа) нулями до формирования целого числа 6-битных групп. Заполнение в конце данных осуществляется как раз с использованием символа «=».

Поскольку входная информация base64 всегда включает целое число октетов, возможны лишь перечисленные ниже случаи:

* размер финального блока кодирования на входе кратен 24 битам, кодированный результат будет содержать целое число 4-символьных групп без заполнения символами «=»;
* размер финального блока кодирования на входе составляет  
  8 битов, выходной блок будет представлять 2 символа, дополненные последовательностью из двух символов заполнения «==»;
* размер финального блока кодирования на входе составляет  
  16 битов, выходной блок будет представлять 3 символа, дополненные символом заполнения «=».

# Практическая часть

1. Создать собственное приложение или воспользоваться *Base64-онлайн-кодировщиком*, с помощью которого конвертировать произвольный документ А на латинице в Б формата base64. В качестве входных данных использовали текстовую строку.

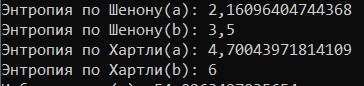
На рисунке 1 показан метод который кодирует сообщение на латинице в формат кодировки base64.

Входная строка: ssssshedko

Base64: c3Nzc3NoZWRrbw==

2. С помощью приложения, созданного в лабораторной работе № 1, получить распределение частотных свойств алфавитов по документам (а) и (б). Вычислить энтропию Хартли и Шеннона, а также избыточность алфавитов. Объяснить полученный результат.

На рисунке 2 показан метод подсчета распределения частотных частиц алфавита. На рисунке 3 показан метод подсчета энтропии хартли и шенона, а так же избыточностьалфавита.



*Рис.2 – Вычисление энтропии по Шенону и Хартли*

3. Написать функцию, которая принимает в качестве аргументов два буфера (*а* и *b*)одинакового размера и возвращает XOR (собственная фамилия (*а*) и имя (*b*); при разной длине меньшую дополнить нулями). Входные аргументы представлять: 1) в кодах ASCII; 2) в кодах base64.

Скриншот выполнения программы представлен на рисунке 5



*Рис.2 – Результат операции XOR*

# Вывод:

В данной лабораторной работе мною были изучены особенности практической реализации операции XOR над данными, представленными в разных форматах. Также я приобрелпрактические навыки трансформацииданных и сопоставление энтропийных свойств используемых приэтом алфавитов.