Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Лабораторная работа №2

**Элементы теории информации. Информации данных в различных кодировках**

Студент: Подобед В.Г.

ФИТ 3 курс 6 группа

Преподаватель: Нистюк О.А.

Минск 2024

**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.

1. **Теоретические сведения**

Из энтропийных оценок (алфавитов и сообщений), полученных в ходе выполнения лабораторной работы № 2, мы выяснили, что энтропия зависит от статических характеристик самих алфавитов и сообщений (вспомним энтропию по Шеннону и по Хартли). Энтропия максимальна при равномерном появлении букв на любом месте сообщения. Для характеристики источника сообщений с различным алфавитом представляет интерес сравнение фактической энтропии источника с максимально возможной. В этом смысле введено понятие избыточности источника сообщений, или избыточности алфавита. Избыточностью алфавита называют уменьшение информационной нагрузки на один символ вследствие разной вероятности и взаимозависимости появления его символов в сообщениях.

Кодирование base64 разработано для представления произвольных последовательностей октетов в форме, позволяющей использовать строчные и прописные буквы. Используется 65-символьное подмножество набора символов US-ASCII, обеспечивающее представление одним печатным символом 6 битов данных (дополнительный 65-й символ используется для обозначения функции специальной обработки). 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

**Задание 1:**

Создать собственное приложение (приветствуется!) или воспользоваться Base64-онлайн-кодировщиком, с помощью которого конвертировать произвольный документ (а) на латинице (можно использовать документ из лабораторной работы № 1) в документ (б) формата base64. В качестве входных данных можно использовать указанный преподавателем вариант из списка:

• входные параметры;

• текстовый файл (\*.txt);

• документ Word (\*.doc);

• документ Word (\*.docx);

• документ PowerPoint (\*.ppt, \*.pptx);

• архив (\*.zip);

• текстовая строка;

• случайное число (от 999999);

• PDF-файл;

• архив (\*.rar);

• архив (\*.7z).

Входные данные в файле А: A Way with Words is a radio show and podcast that features light-hearted conversation about language change, debates, and differences, as well as new words, old sayings, slang, family expressions, word origins and histories, etymology, linguistics, regional dialects, word games and puzzles, grammar, books, literature, writers and writing, and more.

Выходные файлы в файле В: QSBXYXkgd2l0aCBXb3JkcyBpcyBhIHJhZGlvIHNob3cgYW5kIHBvZGNhc3QgdGhhdCBmZWF0dXJlcyBsaWdodC1oZWFydGVkIGNvbnZlcnNhdGlvbiBhYm91dCBsYW5ndWFnZSBjaGFuZ2UsIGRlYmF0ZXMsIGFuZCBkaWZmZXJlbmNlcywgYXMgd2VsbCBhcyBuZXcgd29yZHMsIG9sZCBzYXlpbmdzLCBzbGFuZywgZmFtaWx5IGV4cHJlc3Npb25zLCB3b3JkIG9yaWdpbnMgYW5kIGhpc3RvcmllcywgZXR5bW9sb2d5LCBsaW5ndWlzdGljcywgcmVnaW9uYWwgZGlhbGVjdHMsIHdvcmQgZ2FtZXMgYW5kIHB1enpsZXMsIGdyYW1tYXIsIGJvb2tzLCBsaXRlcmF0dXJlLCB3cml0ZXJzIGFuZCB3cml0aW5nLCBhbmQgbW9yZS4=

**Задание 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 из документа А изображен на рисунке 2.1 и документа В изображен на рисунке 2.2.



Рисунок 2.1



Рисунок 2.2

**Задание 3:**

Записать 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. Что будет результатом операции аXORbXORb?

При написании не использовать стандартные функции языка программирования. Итоговые данные сравнить с результатами использования стандартных функций языка программирования (если они есть).

Результат представлен на рисунке 3.1

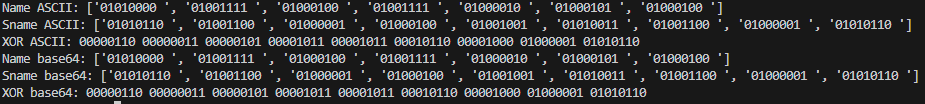


Рисунок 3.1

**Вывод**

В ходе лабораторной работы были закреплены теоретические знания.