Институт информационных технологий

Кафедра: Математическое и программное обеспечение ЭВМ

Дисциплина: Теория Информации

ЛАБОРАТОРНАЯ РАБОТА № 1

Тема: Количественная оценка информации

Выполнил:

студент гр. 1ПИб-02-2оп-23

Кринкин Олег Алексеевич

Проверил:

доцент, к.т.н. Ганичева Оксана Георгиевна

ЗАДАНИЕ

Написать текст, содержащий краткую информацию об одном из основоположников теории информации К. Шенноне. (Рекомендуется взять один из «10 уроков из жизни основоположника теории информации»). Текст должен содержать не менее 7-10 строк, отформатированных по ширине страницы.

Часть 1: использование алфавитного (структурного) подхода

1. Подсчитать количество информации, содержащейся в заданном тексте.
2. Решение задачи оформить в таблице в программе MS Excel.

Считать, что в алфавит символов входят все буквы русского языка.

3) Отчет по работе оформить в программе MS Word, с использованием вставки с решением из программы MS Excel (создать комплексный документ).

Часть 2: использование содержательного(вероятностного) подхода

1. Посчитать количество информации, содержащейся в заданном тексте, используя вероятностный подход.
2. Решение задачи оформить в программе MS Excel.
3. Отчет по работе оформить в программе MS Word, с использованием вставки с решением из программы MS Excel (создать комплексный документ).

ХОД РАБОТЫ

Часть 1: Использование алфавитного (структурного) подхода

1. Составлено следующее сообщение:

Кто такой Клод Шеннон? В математических кругах Шеннон был популярной фигурой. Он считается “отцом информационного века”. В 21 год он опубликовал важную магистерскую работу, которая заложила основу для будущих цифровых компьютеров. И это еще не все. Шеннон является основоположником теории информации, нашедшей применение в современных высокотехнологических системах связи. Он также предложил использовать слово “бит” для обозначения наименьшей единицы информации. Однако Клод Шеннон был не только гениальным теоретиком, но и веселым и изобретательным человеком.

1. Исходное сообщение вставлено в ячейку электронных таблиц B1.
2. Исходя из используемого алфавита, найдена мощность алфавита q = 54 (русский алфавит (мал. буквы) + цифры + спец. символы) и помещена в ячейку B2.
3. В MS Excel при помощи формулы (1.1) найдено количество символов n в тексте. Полученное значение размещено в ячейку B3.

где, – ячейка, содержащая сообщение.

1. По формуле Хартли (1.2) найдено количество информации для одного символа в битах, а затем умножено на количество символов в тексте. Таким образом получено общее количество информации в составленном сообщении.

где, – мощность алфавита

1. Результат расчёта количества информации структурным методом составил 3228,492 бит и представлен на рисунке ниже (рис. 1):



Рис. 1: Результирующая таблица

Часть 2: использование содержательного(вероятностного) подхода.

1. Для работы взято то же сообщение, что и в части 1. Сообщение помещено в ячейку A2 (рис. 2).

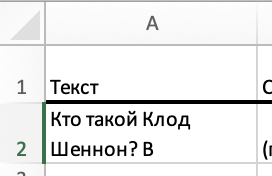


Рис. 2: Вставленный текст

1. Выписаны все символы используемого сообщением алфавита в ячейки A3-A55 (рис. 3).

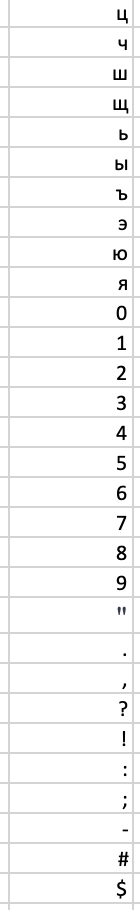
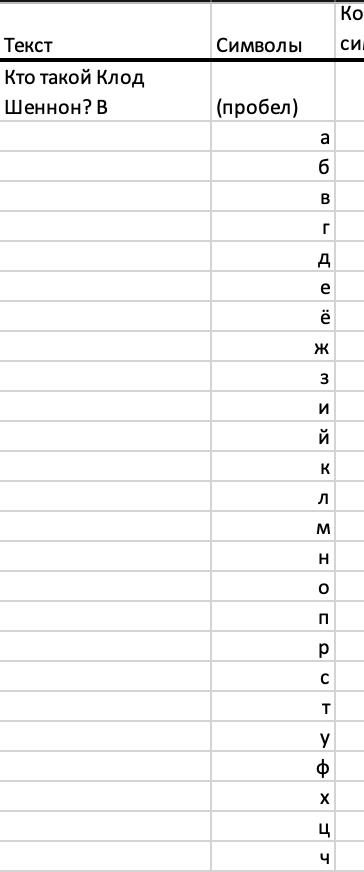


Рис. 3: Выписанный алфавит

1. Для каждого символа в алфавите найдено его количество в тексте путём создания строчки без заданного символа и вычитания длины полученной строки из длины сообщения. Описанные действия выполнены при помощи Excel формулы (2.1):

где, – ячейка с исходным сообщением, – ячейка с символом.

1. Для каждого символа вычислена вероятность появления pi в сообщения при помощи Excel формулы (2.2):

где, – ячейка с исходным сообщением, – сколько раз встречается символ в сообщении.

1. По формуле Шеннона (2.3) вычислено количество информации на один символ в сообщении:

где, – вероятность появления символа в сообщении.

В Excel это действие было реализовано в 2 этапа: для каждой из вероятностей найден член суммы при помощи формулы (2.4):

где, – вероятность появления символа ()

Затем при помощи формулы (2.5) найдена сумма по найденным членам ряда:

где, – ранее вычисленные члены суммы.

1. При помощи Excel формулы (2.6) рассчитано количество информации всего сообщения:

где, – количество информации на один символ, а – длина сообщения.

1. Результат расчёта количества информации статистическим составил 2545,736 бит и представлен на рисунке ниже (рис. 4):



Рис. 4: Результат выполнения статистического метода

РЕЗУЛЬТАТ ВЫПОЛНЕНИЯ РАБОТЫ

Результат нахождения количества информации в сообщении представлен на рисунке ниже (рис. 5):



Рис. 5: Результирующая таблица

ЗАКЛЮЧЕНИЕ

В ходе выполнения лабораторной работы было составлено сообщение длиной в 561 символ. Для этого сообщения при мощности алфавита в 54 символа была выполнена оценка количества информации при помощи двух подходов: структурного и вероятностного.

По результатам оценки количества информации при помощи структурного подхода, объём информации в сообщении составил 3228,492 бит (1).

По результатам оценки количества информации при помощи статистического подхода, объём информации в сообщении составил 2545,736 бит (2).

Таким образом найденное количество информации в сообщении, найденное при помощи структурного подхода оказалось сильно больше, чем найденное при помощи вероятностного подхода (3).

Такое различие объясняется тем, что структурный подход не учитывает точное содержание текста и каждый символ в этом подходе несёт одно и тоже количество информации. Также в данном случае при алфавите из 54 символов некоторые символы не были использованы, но при этом они были включены при расчётах как несущие информацию, хотя это не так. Вероятностный подход, в свою очередь, учитывает точное содержание теста путём нахождения вероятности появления символа. Таким образом при оценке количества информации в сообщении учитываются только те символы, которые действительно несут в себе информацию и содержатся в данном сообщении. Также каждый символ сообщения при вероятностном подходе несёт свой информационный объём что делает этот подход наиболее точным.