

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Сибирский государственный университет телекоммуникаций и
информатики»
(СибГУТИ)

Кафедра прикладной математики и кибернетики

Теория информации
Лабораторная работа №3

Выполнил: студент гр. ИП-711

Мартасов Илья Олегович

Проверила:

Доцент кафедры ПМиК

Мачикина Елена Павловна

Новосибирск 2021 г.

Содержание

Расчет теоретических значений	3
Результаты работы	4
Вывод	5

Расчет теоретических значений

Алфавит источника – 92 символа (для языков C, C++, C#, Python, Lisp):

- прописные и строчные буквы латинского алфавита;
- арабские цифры;
- разделители: , . ; : ? ! ' " | / \ ~ _ ^ () { } [] < > # % & - = + *

Для максимально возможного значения энтропии:

$$P(\text{символ алфавита}) = 1/92$$

$$H = (-1/92 * \log_2[1/92]) * 92 = 6.5872$$

Результаты работы

Язык программирован ия	Максимально возможное значение энтропии	Оценка энтропии (одиночные символы)	Оценка энтропии (частоты пар символов)
C	6.5872	4.88914	3.80512
C++	6.5872	3.95189	3.11008
C#	6.5872	4.33946	3.32293
Python	6.5872	4.3598	3.63511
Lisp	6.5872	3.40941	2,45674

Вывод

В отличии от второй лабораторной работы, здесь разница между оценками энтропии одиночных символов и пар символов значительно больше, поскольку языки программирования более “скудны” на словарный запас, нежели литературные (русский, английский и др.) – в программах часто встречаются одинаковые конструкции (for, while, if...), лексемы и ключевые слов, из-за чего предельная энтропия языков программирования меньше, чем предельная энтропия литературных языков.

Поскольку каждый язык обладает своими конструкциями и лексемами, то из-за этого различаются и оценки энтропий для каждого языка – там, где существует больше ключевых слов, конструкций и лексем, оценка энтропии будет больше, чем в менее “богатых” языках программирования.