

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

Кафедра прикладной математики и кибернетики

Теория информации
Лабораторная работа №2

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 – 32 символа (русский язык, буквы е и ё, ь и ы совпадают, плюс пробел)

Для максимально возможного значения энтропии:

$$P(\text{символ алфавита}) = 1/32$$

$$H = (-1/32 * \log_2[1/32]) * 32 = 5$$

Результаты работы

Название текста	Максимально возможное значение энтропии	Оценка энтропии (одиночные символы)	Оценка энтропии (частоты пар символов)
Безногий и слепой богатыри	5	4.3607	3.91665
Утро помещика	5	4.36693	3.96001
Война и Мир. Том 1	5	4.36214	3.98072
И свет во тьме светит	5	4.34161	3.92354

Вывод

Для рассмотренных текстов русского языка оценки энтропии практически не отличаются между собой. При этом энтропия пар символов каждого текста уже отличается от энтропии, полученной при подсчете частот отдельных символов — это связано с тем, что символы в литературных текстах порождаются в зависимости от его смысла и правил, которые существуют в языке.

При оценке энтропии для пар, троек, четверок символов и так далее, значение оценки будет постепенно уменьшаться, но в какой-то момент изменение между последними двумя рассчитанными оценками будет незначительным. Для русского языка это значение приблизительно равно 2 битам.