

ФЕДЕРАЛЬНОЕ АГЕНТСТВО СВЯЗИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ
БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«СИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ТЕЛЕКОММУНИКАЦИЙ И ИНФОРМАТИКИ»

Теория информации

Практическая работа №2
«Вычисление энтропии Шеннона»

Выполнил:
студент гр. ИП-911
Мироненко К.А.

Проверила:
доцент кафедры ПМиК
Мачикина Е.П.

Новосибирск
2022-2023 уч.год

Результат работы

```
Windows PowerShell
PS C:\Users\User\Desktop\oh shit, here we go again\8 семестр\Information_theory\labs\lab2> python .\lab2.py
Текст: 1984 – George Orwell. Русский язык. Часть первая
Размер алфавита: 43
H0 = log2(43) = 5.426264754702098

Энтропия для 1 символа(ов) подряд: 4.383726570658277
Энтропия для 2 символа(ов) подряд: 3.9946805268364374
PS C:\Users\User\Desktop\oh shit, here we go again\8 семестр\Information_theory\labs\lab2> |
```

Название файла	Размер алфавита	Максимальное возможное значение энтропии	Оценка энтропии (одионые символы)	Оценка энтропии (частоты пар символов)
1984 – George Orwell.	43	≈5.4262647	≈4.383726	≈3.99468

Расчет максимального возможного значения энтропии:

$H = \log(m)$, где m – количество символов в алфавите

Для данного алфавита количество символов равно 43 (32 символов алфавита + 10 цифр + пробел).

$H = \log(43) = 5.426264754702098$

В отличии от первой лабораторной — во второй энтропия для одионых символов и пар сильно отличается, т.к. в художественном тексте у символов больший разброс по их частоте появления, из-за чего неопределённость появления для некоторых букв меньше, чем для других, а некоторые сочетания букв встречаются еще реже, поэтому неопределённость уменьшается еще сильнее.