ФЕДЕРАЛЬНОЕ АГЕНТСТВО СВЯЗИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«СИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ И ИНФОРМАТИКИ»

**Теория информации**

Практическая работа №1-2

«Вычисление энтропии Шеннона»

Выполнил:

студент гр. ИП-911

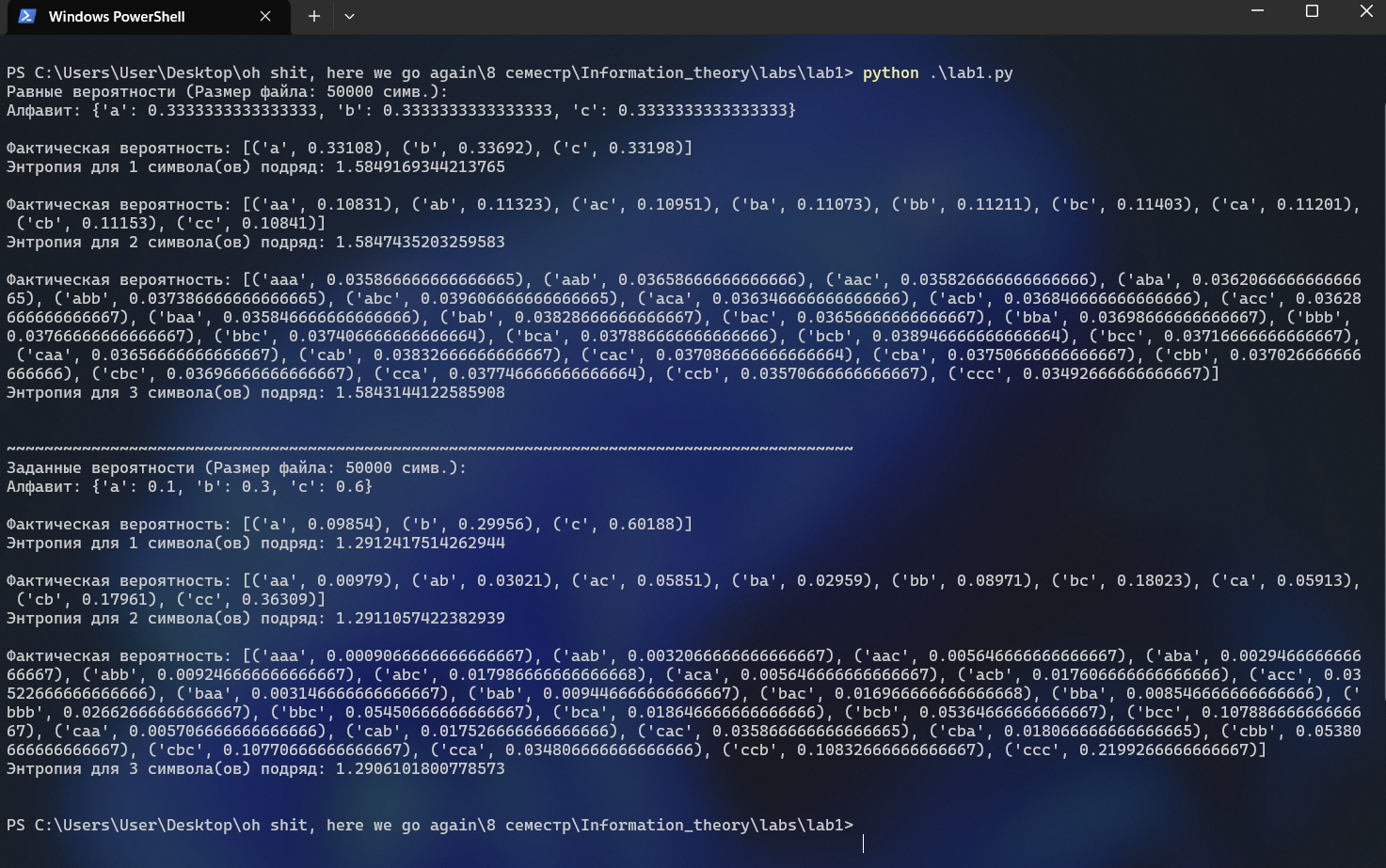
Мироненко К.А.

Проверилa:

доцент кафедры ПМиК

Мачикина Е.П.

# Результат работы первой лабораторной



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Оценка энтропии (частоты отдельных символов) | Теоретическое значение энтропии (отдельные символы) | Оценка энтропии (частоты пар символов) | Теоретическое значение энтропии (для пар символов) |
| Равномерное | ≈ 1.584917 | ≈1.584962 | ≈1.584962 | ≈1.584743 |
| P(0,1; 0,3; 0,6) | ≈1.291241 | ≈1.295462 | ≈1.295461 | ≈1.291105 |

Практические значения были получены в результате работы программы.

Расчет теоретических значений:

1. Теоретическое значение энтропии (отдельные символы):
   1. Для файла, где все цифры генерируются последовательно и независимо с равными вероятностями:

Вероятность выпадения каждого символа равна 1/3;

-p\*log(p) = 0.5283208335737187, где p = 1/3

H1 (0,33; 0,33; 0,33) ≈1.584962

* 1. Для файла, где все цифры последовательности генерируются с заданными вероятностями:

a – 0,1

b – 0,3

c – 0,6

Н1 (0,1; 0,3; 0,6) ≈1.295462

1. Теоретическое значение энтропии (для пар символов):

2.1) Для файла F1:

H2(0.1111; 0.1111; 0.1111; 0.1111; 0.1111; 0.1111; 0.1111; 0.1111; 0.1111) = (0.3522138890491458 \* 9) / 2 ≈1.584743

2.2) Для файла F2:

aa – 0,1 \* 0,1 = 0,01

ab – 0,1 \* 0,3 = 0,03

ac – 0,1 \* 0,6 = 0,06

ba – 0,3 \* 0,1 = 0,03

bb – 0,3 \* 0,3 = 0,09

bc – 0,3 \* 0,6 = 0,18

ca – 0,6 \* 0,1 = 0,06

cb – 0,6 \* 0,3 = 0,18

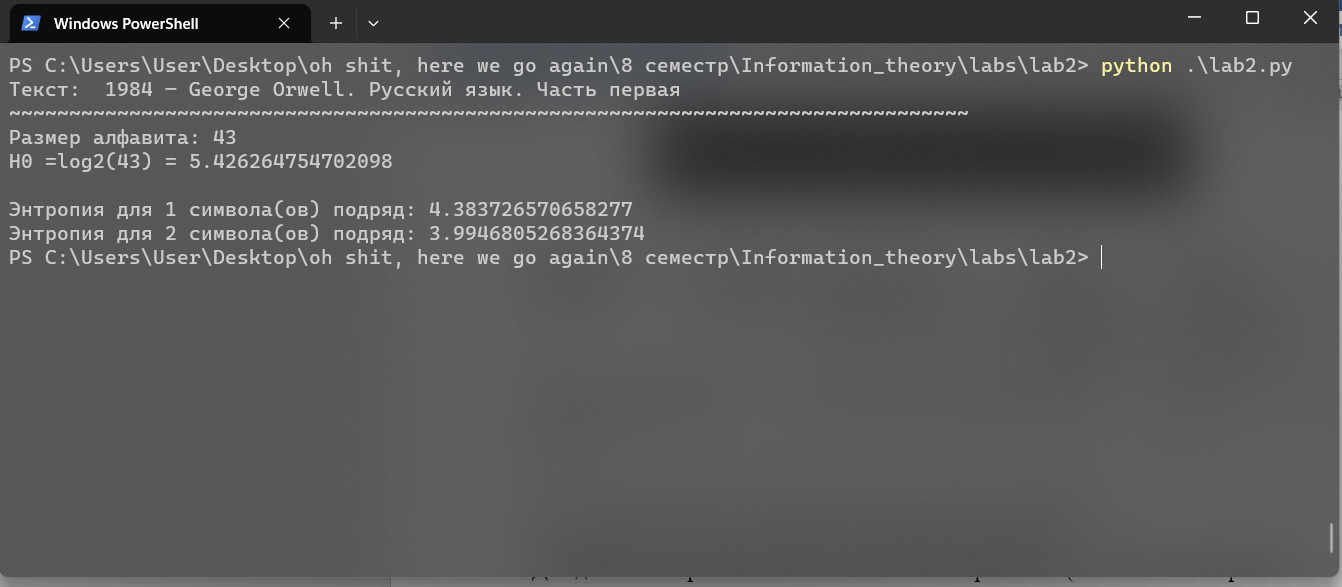
cc – 0,6 \* 0,6 = 0,36

H2 (0,01; 0,03; 0,06; 0,03; 0,09; 0,18; 0,06; 0,18; 0,36) = 2.5909236884766433/ 2 ≈1.291105

**Вывод:**

Сравнив теоретические и практические значения энтропии, можно сказать, что они очень близки. Из этого можно сделать вывод, что программа работает верно.

# Результат работы второй лабораторной



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Название файла | Размер алфавита | Максимальное возможное значение энтропии | Оценка энтропии (одиночные 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

**Вывод:**

В отличии от первой лабораторной — во второй энтропия для одиночных 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.