

Міністерство освіти і науки України Національний технічний університет
України “Київський політехнічний інститут ім. Ігоря Сікорського” Фізико-
технічний інститут

Лабораторна робота №1 з предмету «Криптографія»
«Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Виконав студент 3 курсу

Нечаєв Олексій ФБ-02

Київ - 2023

Мета: засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.

Постановка задачі:

1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H_1 та H_2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H_1 та H_2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 на тому ж тексті, в якому вилючено всі пробіли.
2. За допомогою програми CoolPinkProgram оцінити значення H_{10} , H_{20} та H_{30}
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Результати виконання:

```
H1 без пробілів: 4.348099528385709
R1 без пробілів: 0.12234008182728551
H1 з пробілами: 4.467643407746962
R1 з пробілами: 0.10647131845060753
H2 без пробілів, з перетинами: 4.466786499362919
R2 без пробілів, з перетинами: 0.09838322514633946
H2 з пробілами, з перетинами: 4.649421926530109
R2 з пробілами, з перетинами: 0.0701156146939782
H2 без пробілів, без перетинів: 4.468091761958995
R2 без пробілів, без перетинів: 0.09811975908357173
H2 з пробілами, без перетинів: 4.649701057697397
R2 з пробілами, без перетинів: 0.07005978846052052
```

H10:

[illegible]

$$1.747 < H_{10} < 2.410$$

$$0.6506 > R > 0.518$$

H20:

[illegible]

$$1.964 < \text{H2O} < 2.682$$

$$0.6072 > R > 0.4636$$

Н30:

Лабораторная работа №1

Произвольная часть текста:
гому_каждый_человек_постоянно

Использованные буквы:

Порядок n-граммы:
5 символов
10 символов
15 символов
20 символов
25 символов
30 символов
35 символов
40 символов
45 символов
50 символов

Введенный символ:
Символ по счету:
Номер эксперимента: 52
Поле ввода символов:

Продолжить Другой

Неравенство для энтропии:
1,66849441167849< H < 2,20416163885974
Двоичная таблица угаданных символов:
00100000000000000000000000000000
10000000000000000000000000000000
00000000000000000000000000000000
01000000000000000000000000000000
10000000000000000000000000000000

Вероятности:
q[1]=0,5686274
q[2]=0,1568627
q[3]=0,0784313
q[4]=0
q[5]=0,0196078
q[6]=0
q[7]=0
q[8]=0
q[9]=0
q[10]=0
q[11]=0
q[12]=0
q[13]=0
q[14]=0
q[15]=0
q[16]=0,019607
q[17]=0
q[18]=0
q[19]=0
q[20]=0
q[21]=0
q[22]=0,019607
q[23]=0
q[24]=0,039215
q[25]=0
q[26]=0,039215
q[27]=0
q[28]=0
q[29]=0,019607
q[30]=0,019607
q[31]=0,019607
q[32]=0

Строка состояния:

$$1.669 < H_{30} < 2.204$$
$$0.6662 > R > 0.5592$$

Висновки:

Під час виконання даної роботи я навчився вимірювати частоти символів та біграм у тексті, визначати ентропію. Також, я навчився визначати надлишковість мови.