

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ НАЦІОНАЛЬНИЙ
ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ "КИЇВСЬКИЙ
ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО"
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

КРИПТОГРАФІЯ
КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Виконали:

Студенти групи ФБ-03
Митрофанова М.М. та Мец Є.В.

Київ – 2022

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.

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 1 Мб), де імовірності замінити відповідними частотами. Також одержати значення H^1 та H^2 на тому ж тексті, в якому вилучено всі пробіли.

letters_frequency_no_spaces: Блокнот					letters_frequency_with_spaces: Блокнот				
Файл	Редактирование	Формат	Виглад	Довідка	Файл	Редактирование	Формат	Виглад	Довідка
о - 0.110756					о - 0.162192				
е - 0.084749					е - 0.092792				
а - 0.084186					а - 0.071004				
и - 0.067809					и - 0.078532				
н - 0.064174					н - 0.056811				
т - 0.060555					и - 0.053766				
с - 0.052438					т - 0.050734				
л - 0.049874					с - 0.043933				
р - 0.047218					л - 0.041785				
в - 0.041883					р - 0.03956				
м - 0.034313					в - 0.03509				
к - 0.032182					м - 0.028747				
д - 0.030327					к - 0.026693				
п - 0.02831					д - 0.025408				
у - 0.027785					п - 0.023718				
я - 0.023777					у - 0.023278				
ь - 0.023373					я - 0.019921				
ы - 0.017331					ь - 0.019582				
г - 0.016374					ы - 0.01452				
з - 0.016098					г - 0.013718				
б - 0.014998					з - 0.013487				
ч - 0.014988					б - 0.012566				
ж - 0.011449					ч - 0.012557				
й - 0.010862					ж - 0.009592				
ш - 0.009188					й - 0.0091				
х - 0.008726					ш - 0.007697				
ю - 0.005188					х - 0.007311				
щ - 0.003617					ю - 0.004347				
э - 0.003557					щ - 0.00303				
ц - 0.002301					э - 0.00298				
ф - 0.001613					ц - 0.001928				
					ф - 0.001352				

1) bigrams_no_spaces

[illegible]

Посилання на папку з таблицками:

https://drive.google.com/drive/folders/1pEcjDZcUFTM9kulQN3bzGm1Kw1_mdcdh?usp=sharing

- Ентропія

```
1 Letters with whitespaces entropy: 4.371059 ; redundancy: 0.125788
2 Letters without whitespaces entropy: 4.453914 ; redundancy: 0.100982
3 Bigrams with whitespaces with crossing entropy: 3.9558 ; redundancy: 0.20884
4 Bigrams with whitespaces without crossing entropy: 3.951953 ; redundancy: 0.209609
5 Bigrams without whitespaces with crossing entropy: 4.126059 ; redundancy: 0.167159
6 Bigrams without whitespaces without crossing entropy: 4.12453 ; redundancy: 0.167467
```

2. За допомогою програми CoolPinkProgram оцінити значення H^{10} , H^{20} , H^{30} .
 H^{10} :

Лабораторная работа №1

Произвольная часть текста:
нны_ка_

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: 55

Неравенство для энтропии:
 $2,81001450661552 < H < 3,519642605788$

Двоичная таблица угаданных символов:

Поле ввода символов:

Продолжить Другой

Вероятности:

Строка состояния:

$$2,8100 < H^{10} < 3,5196$$

H^{20} :

Лабораторная работа №1

Произвольная часть текста:
исля_но_понятие_о_д

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: 55

Неравенство для энтропии:
 $1,83390965865442 < H < 2,512482675750$

Двоичная таблица угаданных символов:

Поле ввода символов:

Продолжить Другой

Вероятности:

Строка состояния:

$$1,8339 < H^{20} < 2,51248$$

H^{30} :

Лабораторная работа №1

Произвольная часть текста:
я_труд_сравнить_учения_о_моря

Использованные буквы:

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

Введенный символ:
Символ по счету:
Номер эксперимента: 56

Неравенство для энтропии:
 $1,90455438670184 < H < 2,590970979438$

Двоичная таблица угаданных символов:

1000000000000000000000000000000000
0100000000000000000000000000000000
0000000000000000000000000000000001
1000000000000000000000000000000000
0000000001000000000000000000000000

Вероятности:

q[1] = 0,045454545454545
q[2] = 0,109090909090909
q[3] = 0
q[4] = 0,018181818181818
q[5] = 0,036363636363636
q[6] = 0,054545454545454
q[7] = 0,018181818181818
q[8] = 0,018181818181818
q[9] = 0,018181818181818
q[10] = 0
q[11] = 0,018181818181818
q[12] = 0
q[13] = 0
q[14] = 0,036363636363636
q[15] = 0
q[16] = 0
q[17] = 0
q[18] = 0,018181818181818
q[19] = 0
q[20] = 0
q[21] = 0,018181818181818
q[22] = 0
q[23] = 0
q[24] = 0,036363636363636
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0
q[29] = 0,036363636363636
q[30] = 0
q[31] = 0
q[32] = 0,018181818181818

Строка состояния:

$$1,90455 < H^{20} < 2,59097$$

3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

$$R = 1 - \frac{H_{\infty}}{H_0}$$

H^{10} :

$$1) R = 1 - \frac{2,81001450661552}{5} \approx 0,4379$$

$$2) R = 1 - \frac{3,519642605788}{5} \approx 0,2960$$

H^{20} :

$$3) R = 1 - \frac{1,83390965865442}{5} \approx 0,6332$$

$$4) R = 1 - \frac{2,512482675750}{5} \approx 0,4975$$

H^{30} :

$$5) R = 1 - \frac{1,90455438670184}{5} \approx 0,62$$

$$6) R = 1 - \frac{2,590970979438}{5} \approx 0,482$$

Труднощі:

- Вивід матриці біграм

ВИСНОВКИ

Під час виконання комп'ютерного практикуму ми ознайомились з поняттям ентропії на символ джерела та його надлишковості, написали програми для підрахунку частот букв і частот біграм в тексті, попрацювали з CoolPinkProgram, де оцінювали та порівнювали H^{10} , H^{20} , H^{30} . В результаті експериментів ми отримали значення ентропії та надлишковості власного джерела, які наближаються до значень, отриманих з прикладеної програми.