Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Фізико-технічний інститут

КРИПТОГРАФІЯ

Виконав:
Студент гр. ФБ-31 Моісеєнко Д.Ю.

Київ-2026

Комп’ютерний практикум №1. Експериментальна оцінка ентропії на символ джерела відкритого тексту

Мета роботи: 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

4 варіант

***Порядок виконання роботи**:*

0. Уважно прочитати методичні вказівки до виконання комп’ютерного практикуму.

1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H1 та H2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H1 та H2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H1 та H2 на тому ж тексті, в якому вилучено всі пробіли.

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

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

Хід роботи:

0.

```
PS D:\3 курс 5 семестр 2025(академічна різниця)\Криптографія\Криптографія загальна 2025\crypto25-26\lab1\moiseienko_fb-31_cp1> python lab1.py
Обробка файлу: Брати Карамазови Федір Достоєвський.txt
Успіх
Створений файл з пробілами: with_spaces.txt (символів: 621970)
Створений файл без пробілів: no_spaces.txt (символів: 515917)
```

1. **Результати розрахунків ентропії:**

Текст	Параметри	Значення
З пробілами	Ентропія *H1*	4.36
Без пробілів	Ентропія *H2*	4.46

З пробілами	Ентропія *H2* з перетинами	3.95
Без пробілів	Ентропія *H2* з перетинами	4.14
З пробілами	Ентропія *H2* без перетинів	3.952
Без пробілів	Ентропія *H2* без перетинів	4.143

Також додаю скріншоти з літерами та біграми з Excel-file:

Frequencies_bigrams:

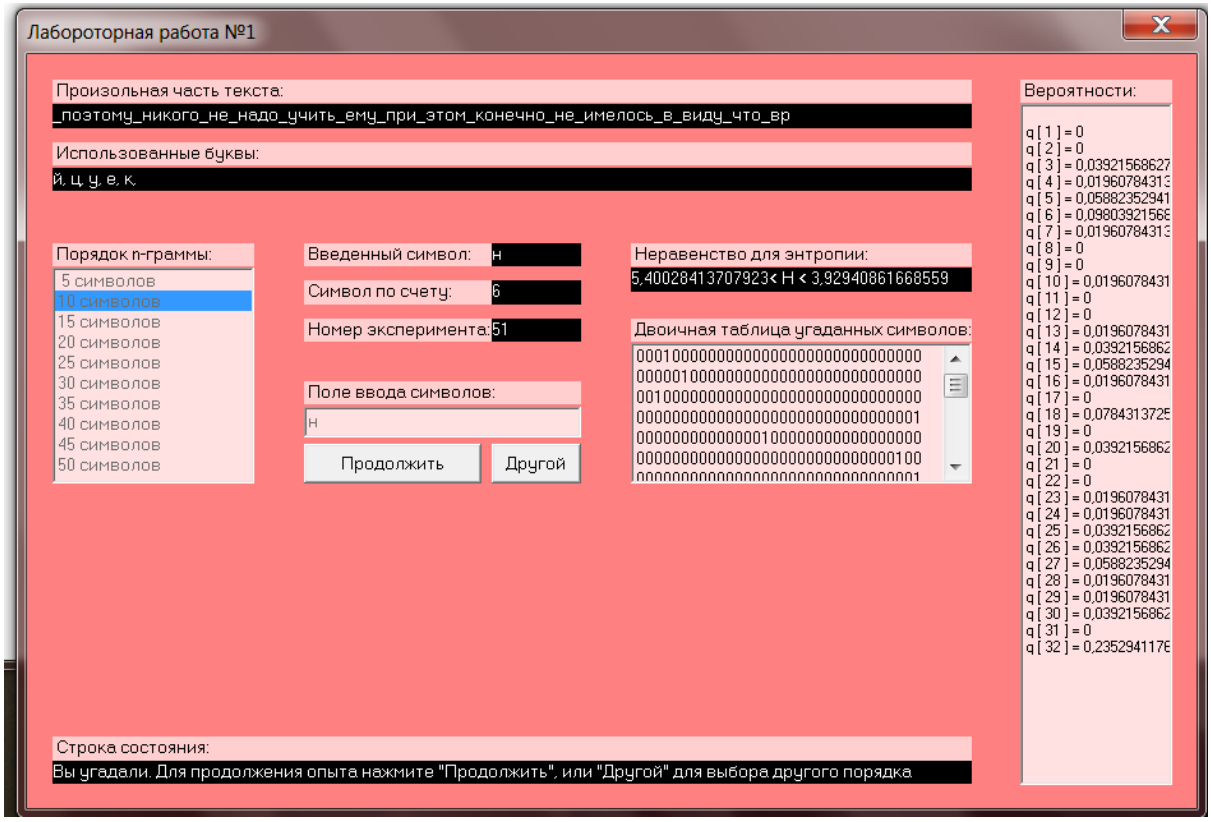
	Елемент	Кількість	Частота
2	то	4719	0.018293675714651222
3	ст	3256	0.012622209817102008
4	не	3239	0.01255630761596849
5	ов	3121	0.012098868808100544
6	но	3031	0.011749974802099567
7	на	2937	0.01138557439583188
8	го	2779	0.01077307158529683
9	ен	2544	0.009862070569627613
10	от	2536	0.009831057769094194
11	по	2474	0.009590708564960188
12	ос	2455	0.009517053163693315
13	ко	2423	0.009393001961559634
14	ал	2419	0.009377495561292923
15	он	2403	0.009315469960226083
16	во	2167	0.008400592344490188
17	ни	2107	0.008167996340489538
18	ка	2095	0.008121477139689406
19	ро	2090	0.008102094139356019
20	пр	2068	0.008016808937889114
21	ес	2042	0.007916017336155498
22	ет	2015	0.007811349134355205
23	ть	1991	0.007718310732754944
24	ра	1984	0.007691174532288202
25	ли	1919	0.007439195527954163
26	ор	1891	0.007330650726087192
27	ер	1879	0.007284131525287062
28	та	1877	0.007276378325153707
29	ак	1860	0.007210476124020189 5
30	ит	1841	0.007136820722753317
31	те	1803	0.00698950992021957

frequencies_letters:

	A	B	C	D	E	F
1	Елемент	Кількість	Частота			
2	о	58211	0.1128301645419709			
3	е	46439	0.09001254077690791			
4	а	40138	0.07779933593969572			
5	и	34656	0.0671735957528052			
6	т	34166	0.06622383057739908			
7	н	31944	0.06191693625137377			
8	с	27555	0.053409753894521794			
9	в	24349	0.04719557603257889			
10	л	22815	0.04422222954467482			
11	р	20897	0.04050457728665657			
12	м	17101	0.03314680462167364			
13	д	16610	0.03219510114999118			
14	к	16523	0.032026469373949686			
15	у	14635	0.02836696600422161			
16	п	12753	0.024719092412151568			
17	ь	10946	0.02121659104080695			
18	я	10459	0.020272640754229847			
19	г	10000	0.019382962763390237			
20	ч	9443	0.0183033317374694			
21	б	9368	0.018157959516743972			
22	ы	8801	0.017058945528059748			
23	з	7746	0.015014042956522077			
24	ж	5646	0.010943620776210127			
25	ш	5333	0.010336934041716012			
26	й	5306	0.010284600042254859			
27	х	4164	0.008071065694675694			
28	ю	3598	0.006973990002267807			
29	ц	1966	0.00381069047928252			
30	э	1852	0.0035897247037798715			
31	щ	1519	0.002944272043758977			
32	ф	978	0.001895653758259565			

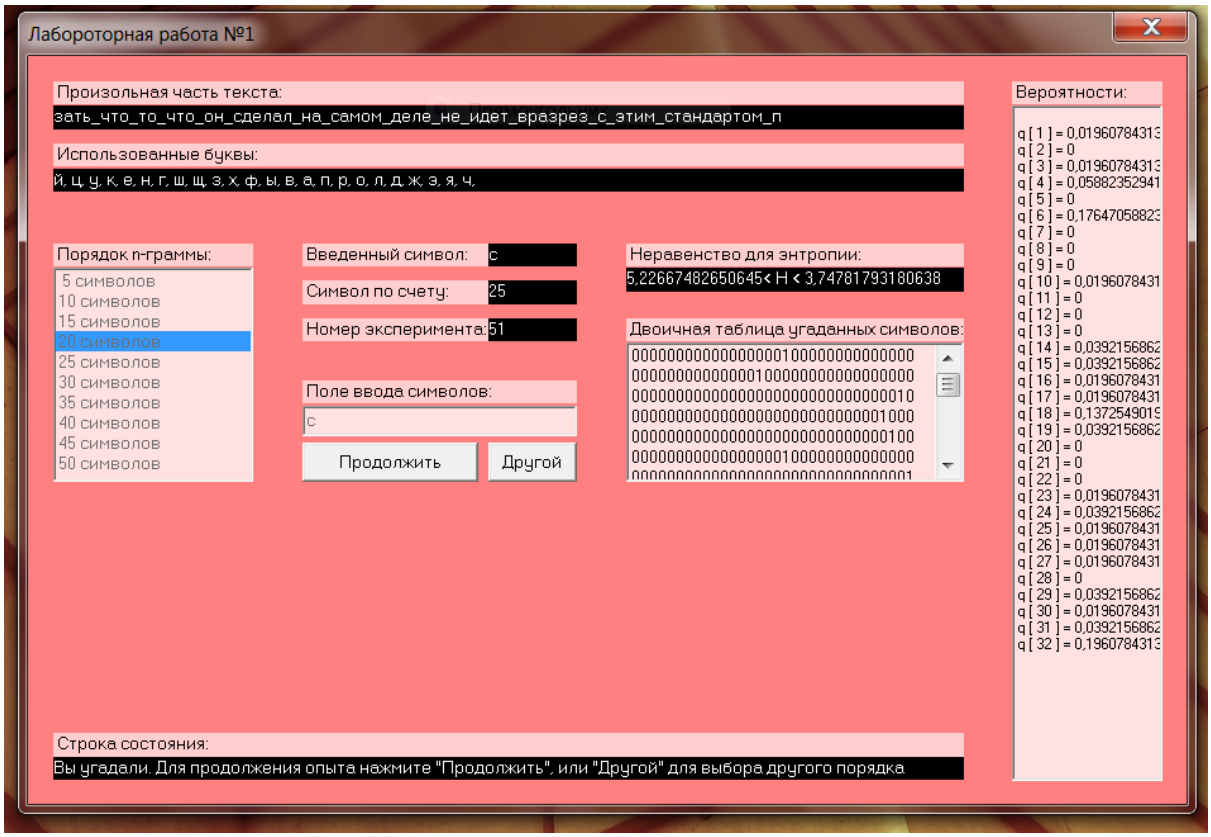
2. **Експериментальна рожева програмка(CoolPinkProgram):**

H^{10}**:**


Лабороторная работа №1
Произольная часть текста:
_поэтому_никого_не_надо_учить_ему_при_этом_конечно_не_имелось_в_виду_что_вр
Использованные буквы:
й, ц, у, е, к,
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: н
Символ по счету: 6
Номер эксперимента: 51
Поле ввода символов:
н
Продолжить
Другой
Неравенство для энтропии:
5,40028413707923< H < 3,92940861668559
Двоичная таблица угаданных символов:
Вероятности:
q[1] = 0
q[2] = 0
q[3] = 0,03921568627
q[4] = 0,01960784313
q[5] = 0,05882352941
q[6] = 0,09803921568
q[7] = 0,01960784313
q[8] = 0
q[9] = 0
q[10] = 0,0196078431
q[11] = 0
q[12] = 0
q[13] = 0,0196078431
q[14] = 0,0392156862
q[15] = 0,0588235294
q[16] = 0,0196078431
q[17] = 0
q[18] = 0,0784313725
q[19] = 0
q[20] = 0,0392156862
q[21] = 0
q[22] = 0
q[23] = 0,0196078431
q[24] = 0,0196078431
q[25] = 0,0392156862
q[26] = 0,0392156862
q[27] = 0,0588235294
q[28] = 0,0196078431
q[29] = 0,0196078431
q[30] = 0,0392156862
q[31] = 0
q[32] = 0,2352941176
Строка состояния:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

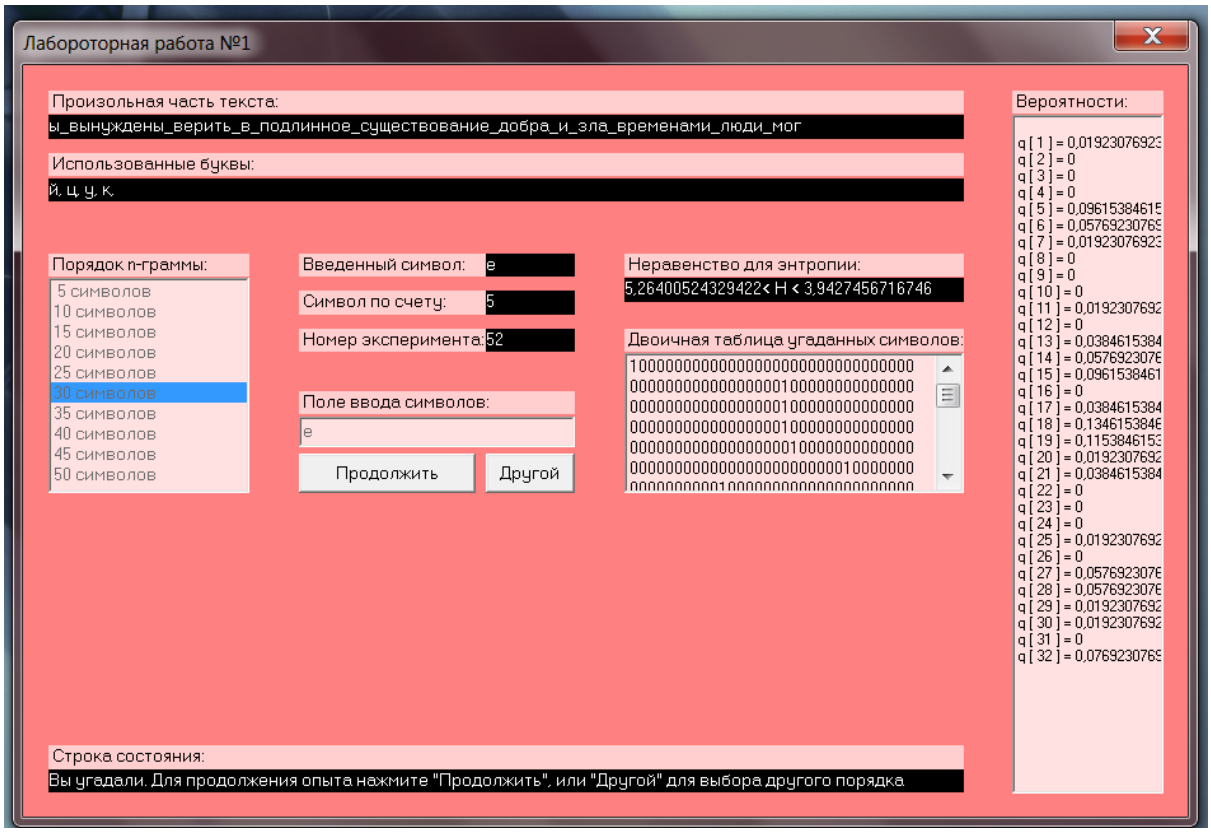
H < 3.92

H^{20}:


Лабороторная работа №1
Произольная часть текста:
зать_что_то_что_он_сделал_на_самом_деле_не_идет_вразрез_с_этим_стандартом_п
Использованные буквы:
й, ц, у, к, е, н, г, ш, щ, з, х, ф, ы, в, а, п, р, о, л, д, ж, э, я, ч,
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: с
Символ по счету: 25
Номер эксперимента: 51
Поле ввода символов:
с
Продолжить
Другой
Неравенство для энтропии:
5,226674826506 45< H < 3,74781793180638
Двоичная таблица угаданных символов:
Вероятности:
q[1] = 0,0196078431
q[2] = 0
q[3] = 0,0196078431
q[4] = 0,05882352941
q[5] = 0
q[6] = 0,1764705882
q[7] = 0
q[8] = 0
q[9] = 0
q[10] = 0,0196078431
q[11] = 0
q[12] = 0
q[13] = 0
q[14] = 0,0392156862
q[15] = 0,0392156862
q[16] = 0,0196078431
q[17] = 0,0196078431
q[18] = 0,1372549019
q[19] = 0,0392156862
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0,0196078431
q[24] = 0,0392156862
q[25] = 0,0196078431
q[26] = 0,0196078431
q[27] = 0,0196078431
q[28] = 0
q[29] = 0,0392156862
q[30] = 0,0196078431
q[31] = 0,0392156862
q[32] = 0,1960784313
Строка состояния:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

H < 3.74

H^{30}:



H < 3.94

Напишу свій опис труднощів з яким стикався під час початку експерименту. Справа в тому, що на Windows 11 не дуже коректно відображалася програма і букви були у вигляді “карлючки”. Тому мною було найкраще рішення спробувати на більш старіші ноутбуки(Windows7) і запустити Рожеву програму. В цілому вона спрацювало, напис програми став зрозумілим і в цілому зміг провести експериментальну частину даної лабораторної роботи.

3. Розрахунок надлишковості:

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

Зазначу, що $H_0 = log_2 32 = 5$.

N	Ентропія H_n	Надлишковість R
n = 1	4.46	**10.76 %**
n = 2	4.14	**17.14 %**
H^{10}	3.92	**21.42 %**
H^{20}	3.74	**25.06 %**
H^{30}	3.94	**21.16 %**

4. Значення надлишковості різних мов:

Надмірність природних мов [ред . редагувати код]

Розмір надмірності різних мов світу коливається не більше 70—80% [5] . У всіх мовах всіх рівнях присутні надлишкові елементи. Надмірність у мові невипадкова: її функція — полегшити комунікацію за несприятливих умов передачі. Надмірність є системою попередження можливих помилок [6] .

Для англійського тексту, що складається з 26 літер, $H_{max} = \log_2 26 = 4.7$. Шенноном було встановлено, що ентропія англійського тексту при $n = 100$ дорівнює 0.6-1.3 біт / символ [7] , що може бути прийнято за ентропію англійського тексту [8] . Отже, надмірність англійського тексту становить 72—87%.

Також з допомогою експериментальних оцінок було визначено ентропії інших мов. У таблиці представлені ентропії російської та французької мов, а також їх надмірності для різних типів тексту [3] .

Тип тексту	Ентропія російського тексту	Ентропія французького тексту	Надмірність російського тексту, %	Надмірність французького тексту, %
Загалом	1,37	1,40	72,6	70,6
Розмовний текст	1,40	1,50	72,0	68,4
Літературний текст	1,19	1,38	76,2	71,0
Діловий текст	0,83	1,22	83,4	74,4

Висновок:

У ході виконання лабораторної роботи було проведено експериментальну оцінку ентропії на символ джерела відкритого тексту. Утім, що при збільшенні порядку *N* значення ентропії закономірно зменшується, а надлишковість мови навпаки зростає такий як є підтвердженням сильних статистичних зв’язків між символами.