

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

**Криптографія**  
Лабораторна робота №1

Виконала:  
Студентка 3 курсу  
Групи ФБ-25 Ляшенко Аліна

Біла Церква -2024

**Тема:** експериментальна оцінка ентропії на символ джерела відкритого тексту.

**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.

## Хід 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  $H_1$  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 вилучено всі пробіли.

- Підраховану частоту букв, запишу у окремі текстові файли. Підрахунок частоти букв у тексті з пробілами у letter\_with\_gaps.txt. А частоту букв у тексті без пробілів - letter\_without\_gaps.txt.

letter_with_gaps.txt	letter_without_gaps.txt
1 а: 0.064802	1 а: 0.064862
2 б: 0.012173	2 б: 0.012167
3 в: 0.044395	3 в: 0.044518
4 г: 0.015395	4 г: 0.015183
5 д: 0.025778	5 д: 0.025822
6 е: 0.087000	6 е: 0.086774
7 ж: 0.005012	7 ж: 0.005002
8 з: 0.015037	8 з: 0.015064
9 и: 0.113493	9 и: 0.114004
10 й: 0.017901	10 й: 0.017982
11 к: 0.045111	11 к: 0.045312
12 л: 0.023630	12 л: 0.023718
13 м: 0.023672	13 м: 0.023658
14 н: 0.070889	14 н: 0.071094
15 о: 0.087716	15 о: 0.088084
16 п: 0.025420	16 п: 0.025464
17 р: 0.053345	17 р: 0.053370
18 с: 0.057642	18 с: 0.057955
19 т: 0.070172	19 т: 0.070002
20 у: 0.034370	20 у: 0.034436
21 ф: 0.003580	21 ф: 0.003573
22 х: 0.018975	22 х: 0.018299
23 ц: 0.007518	23 ц: 0.007542
24 ч: 0.014679	24 ч: 0.014747
25 ш: 0.004654	25 ш: 0.004644
26 щ: 0.003222	26 щ: 0.003235
27 ы: 0.022913	27 ы: 0.022289
28 ь: 0.012173	28 ь: 0.012167
29 э: 0.001074	29 э: 0.001072
30 ю: 0.002506	30 ю: 0.002501
31 я: 0.015753	31 я: 0.015461

- Підраховану частоту біграм, запишу у окремі таблиці. Підрахунок частоти біграм у тексті з пробілами у bigram\_with\_gap.xlsx. А частоту біграм у тексті без пробілів - bigram\_without\_gap.xlsx.

bigram\_with\_gap.xlsx

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у
а	0	0.00179015	0.00286424	0.00071606	0.00143212	0.00250621	0.00107409	0.00286424	0.00429636	0	0.00358030	0.00286424	0.00393833	0.00930879	0.00035803	0.00071606	0.00429636	0.00322227	0.00465439	0.00501242
б	0.00107409	0	0	0	0	0.00143212	0	0	0.00035803	0	0	0.00107409	0	0.00179015	0.00286424	0	0.00143212	0.00071606	0	0.00071606
в	0.00537045	0	0.00035803	0.00107409	0	0.01038288	0	0.00035803	0.00429636	0	0.00179015	0.00179015	0.00071606	0.00286424	0.00286424	0.00071606	0.00214818	0.00286424	0.00107409	0.00143212
г	0.00143212	0	0	0	0	0.00035803	0	0	0.00143212	0	0.00035803	0	0	0	0.01109894	0	0.00071606	0	0	0
д	0.00537045	0.00035803	0.00071606	0.00071606	0	0.00393833	0	0	0.00250621	0	0	0	0	0.00358030	0.00286424	0	0.00143212	0	0	0.00179015
е	0	0.00179015	0.00429636	0.00214818	0.00537045	0.00250621	0.00286424	0.00107409	0.00143212	0.00465439	0.00179015	0.00286424	0.00286424	0.01038288	0.00035803	0.00214818	0.00895076	0.00787666	0.01396318	0.00107409
ж	0	0	0	0	0.00214818	0.00107409	0	0	0.00071606	0	0	0	0	0.00071606	0	0	0	0	0	0
з	0.00537045	0	0.00107409	0	0.00143212	0.00107409	0	0	0.00035803	0	0.00071606	0.00035803	0	0.00107409	0.00214818	0	0.00071606	0.00035803	0	0.00035803
и	0.00143212	0.00107409	0.00787666	0.00322227	0.00322227	0.00501242	0.00035803	0.00465439	0.01074091	0.00680257	0.01034076	0.00107409	0.00718166	0.01074091	0.00286424	0.00107409	0.00179015	0.00572848	0.01253106	0.00250621
й	0.00035803	0	0.00035803	0.00071606	0.00035803	0.00035803	0	0	0.00250621	0	0.00071606	0	0.00071606	0.00035803	0.00107409	0.00250621	0	0.00214818	0.00107409	0.00143212
к	0.00680257	0	0	0	0	0.00035803	0.00035803	0	0.01109894	0	0.00071606	0.00035803	0	0.00035803	0.01038288	0.00465439	0.00465439	0.00035803	0.00179015	0.00322227
л	0.00107409	0	0	0	0	0.00429636	0	0	0.00537045	0	0	0	0	0	0.00250621	0	0	0	0.00035803	0.00143212
м	0.00179015	0	0.00035803	0	0	0.00358030	0.00035803	0	0.00823469	0	0.00073712	0.00035803	0.00109515	0.00035803	0.00179015	0.00179015	0	0.00071606	0	0.00107409
н	0.0096682	0.00035803	0.00035803	0.00107409	0.00071606	0.00537045	0	0	0.01933364	0	0	0	0.00035803	0.00358030	0.00787666	0.00035803	0	0.00250621	0.00429636	0.00035803
о	0	0.00429636	0.01682742	0.00358030	0.00680257	0.00250621	0	0.00214818	0.00143212	0.00250621	0.00322227	0.00680257	0.00286424	0.00286424	0.00179015	0.00214818	0.01611136	0.00358030	0.00537045	0
п	0.00179015	0	0	0	0.00179015	0	0	0	0.00501242	0	0	0.00071606	0	0.00107409	0.00787666	0	0.00393833	0	0	0.00143212
р	0.01360515	0	0	0.00107409	0	0.00537045	0	0	0.00322227	0	0.00107409	0	0.00071606	0.00179015	0.00823469	0.00107409	0.00035803	0.00895076	0.00107409	0.00250621
с	0.00107409	0	0.00035803	0.00035803	0	0.00143212	0	0	0.00930879	0	0.01074091	0.00179015	0	0.00071606	0.00358030	0.00286424	0.00035803	0.00179015	0.01467924	0.00250621
т	0.00429636	0	0.00465439	0.00035803	0.00035803	0.01790152	0	0	0.00751863	0	0.00071606	0	0.00035803	0.00107409	0.01396318	0.00035803	0.00358030	0.00608651	0	0.00465439
у	0	0.00071606	0	0	0.00322227	0.00035803	0	0.00179015	0	0	0.00501242	0.00286424	0	0.00537045	0	0.00143212	0.00071606	0.00143212	0.00214818	0
ф	0.00143212	0	0	0	0	0.00143212	0	0	0.00035803	0	0	0	0	0	0.00035803	0	0	0	0	0
х	0	0.00107409	0.00143212	0	0	0.00035803	0	0.00143212	0	0.00143212	0	0	0.00035803	0.00035803	0.00644454	0.00107409	0.00035803	0.00071606	0.00107409	0.00035803
ц	0.00035803	0	0	0	0	0.00179015	0	0	0.00501242	0	0	0	0	0	0	0	0	0	0	0
ч	0.00214818	0	0.00035803	0	0	0.00644454	0	0	0.00107409	0	0	0	0	0.00286424	0	0	0	0.00071606	0	0
ш	0	0	0	0	0	0.00107409	0	0	0.00322227	0	0	0	0	0	0.00035803	0	0	0	0	0
щ	0.00035803	0	0	0	0	0.00143212	0	0	0.00107409	0	0	0	0	0	0	0	0	0	0	0.00035803
ы	0	0	0.00107409	0	0.00071606	0.00107409	0	0	0.00107409	0.00393833	0.00143212	0.00035803	0.00143212	0.00035803	0	0.00143212	0.00035803	0.00358030	0	0
ь	0	0	0.00035803	0	0	0	0	0.00035803	0.00035803	0	0.00179015	0	0	0.00286424	0.00035803	0.00035803	0	0.00250621	0.00214818	0
э	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00107409	0
ю	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00035803	0	0.00143212	0

bigram\_without\_gap.xlsx

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у
а	0	0.00178631	0.00287795	0.00073437	0.00142905	0.00250084	0.00107179	0.00285810	0.00432685	0	0.00357263	0.00289780	0.00392989	0.00934839	0.00035726	0.00071452	0.00428716	0.00321537	0.00464442	0.00500168
б	0.00107179	0	0	0	0	0.00142905	0	0	0.00035726	0	0	0.00107179	0	0.00178631	0.00287795	0	0.00142905	0.00071452	0	0.00071452
в	0.00539864	0	0.00035726	0.00075422	0	0.01038048	0	0.00035726	0.00434670	0	0.00178631	0.00178631	0.00071452	0.00285810	0.00323521	0.00071452	0.00214358	0.00289780	0.00107179	0.00142905
г	0.00142905	0	0	0	0	0.00035726	0	0	0.00142905	0	0.00035726	0	0	0	0.01089653	0	0.00071452	0	0	0
д	0.00537879	0.00035726	0.00071452	0.00073437	0	0.00392989	0	0	0.00250084	0	0	0	0	0.00357263	0.00285810	0	0.00142905	0	0	0.00180616
е	0	0.00178631	0.00432685	0.00214358	0.00535895	0.00252069	0.00285810	0.00109163	0.00142905	0.00464442	0.00178631	0.00285810	0.00285810	0.01038048	0.00035726	0.00214358	0.00859416	0.00793918	0.01397296	0.00107179
ж	0	0	0	0	0.00214358	0.00107179	0	0	0.00071452	0	0	0	0	0.00071452	0	0	0	0	0	0
з	0.00539864	0	0.00107179	0	0.00142905	0.00107179	0	0	0.00035726	0	0.00071452	0.00035726	0	0.00109163	0.00214358	0	0.00071452	0.00035726	0	0.00035726
и	0.00144890	0.00107179	0.00787964	0.00325506	0.00321537	0.00539864	0.00035726	0.00466427	0.01075759	0.00688724	0.01038048	0.00109163	0.00722465	0.01079729	0.00287795	0.00107179	0.00214358	0.00611317	0.01224619	0.00252069
й	0.00035726	0	0.00035726	0.00071452	0.00035726	0.00035726	0	0	0.00254053	0	0.00071452	0	0.00071452	0.00035726	0.00107179	0.00254053	0	0.00214358	0.00109163	0.00144890
к	0.00678800	0	0	0	0	0.00035726	0.00035726	0	0.01121409	0	0.00071452	0.00035726	0	0.00035726	0.01045987	0.00466427	0.00468411	0.00035726	0.00178631	0.00321537
л	0.00107179	0	0	0	0	0.00430700	0	0.00541849	0	0	0	0	0	0	0.00254053	0	0	0	0.00035726	0.00142905
м	0.00178631	0	0.00035726	0	0	0.00359248	0.00035726	0	0.00823690	0	0.00073437	0.00035726	0.00107179	0.00035726	0.00180616	0.00178631	0	0.00071452	0	0.00107179
н	0.00968580	0.00035726	0.00073437	0.00107179	0.00071452	0.00537879	0	0	0.01943115	0	0	0	0.00035726	0.00357263	0.00787964	0.00037711	0	0.00254053	0.00428716	0.00035726
о	0	0.00428716	0.01687076	0.00361232	0.00688724	0.00250084	0	0.00214358	0.00142905	0.00250084	0.00325506	0.00686739	0.00285810	0.00289780	0.00180616	0.00214358	0.01615624	0.00361232	0.00539864	0
п	0.00178631	0	0	0	0	0.00180616	0	0	0.00502153	0	0	0.00071452	0	0.00107179	0.00791933	0	0.00392989	0	0	0.00142905
р	0.01361570	0	0	0.00107179	0	0.00535895	0	0	0.00321537	0	0.00107179	0	0.00071452	0.00178631	0.00821705	0.00107179	0.00035726	0.00899112	0.00107179	0.00250084
с	0.00107179	0	0.00035726	0.00037711	0	0.00142905	0	0	0.00936823	0	0.01085683	0.00178631	0	0.00073437	0.00357263	0.00287795	0.00035726	0.00178631	0.01474703	0.00252069
т	0.00428716	0	0.00466427	0.00035726	0.00035726	0.01722803	0	0	0.00758192	0	0.00071452	0	0.00035726	0.00107179	0.01397296	0.00035726	0.00392989	0.00609332	0	0.00470396
у	0	0.00071452	0	0	0.00321537	0.00035726	0	0.00180616	0	0	0.00504138	0.00285810	0	0.00537879	0	0.00142905	0.00071452	0.00142905	0.00218327	0
ф	0.00142905	0	0	0	0	0.00142905	0	0	0.00035726	0	0	0	0	0	0.00035726	0	0	0	0	0
х	0	0.00109163	0.00142905	0	0	0.00035726	0	0.00142905	0.00142905	0	0	0.00035726	0.00035726	0.00649028	0.00107179	0.00035726	0.00035726	0.00071452	0.00035726	0.00214358
ц	0.00035726	0	0	0	0	0.00178631	0	0	0.00539864	0	0	0	0	0	0	0	0	0	0	0
ч	0.00214358	0	0.00037711	0	0	0.00649028	0	0	0.00107179	0	0	0	0	0	0.00285810	0	0	0.00073437	0	0
ш	0	0	0	0	0	0.00107179	0	0	0.00321537	0	0	0	0	0	0.00035726	0	0	0	0	0
щ	0.00035726	0	0	0	0	0.00144890	0	0	0.00107179	0	0	0	0	0	0	0	0	0	0	0.00035726
ы	0	0	0.00071452	0	0.00071452	0.00107179	0	0	0.00071452	0.00394974	0.00144890	0.00035726	0.00142905	0.00037711	0	0.00142905	0.00035726	0.00398944	0	0
ь	0	0	0.00035726	0	0	0	0	0.00035726	0.00035726	0	0.00178631	0	0	0.00287795	0.00035726	0.00035726	0	0.00250084	0.00214358	0
э	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00107179	0
ю	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00035726	0	0.00142905	0

H1: 4.398592642298493  
H2: 3.918895792949149  
H1 (no gap): 4.395720245359787  
H2 (no gap): 3.9172223889227653

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

$$- 1,4368 < H^{(10)} < 2,2047$$

Лабораторная работа №1

Произвольная часть текста:  
что, время

Использованные буквы:

Порядок n-граммы:  
5 символов

Введенный символ:  
Символ по счету:  
Номер эксперимента: 50

Неравенство для энтропии:  
 $1,43680132713866 < H < 2,20471268847602$

Двоичная таблица угаданных символов:

Вероятности:

q[1]	= 0,6326530
q[2]	= 0,0612244
q[3]	= 0,0612244
q[4]	= 0,0408163
q[5]	= 0,0204081
q[6]	= 0,0408163
q[7]	= 0,0204081
q[8]	= 0,0204081
q[9]	= 0
q[10]	= 0
q[11]	= 0
q[12]	= 0
q[13]	= 0
q[14]	= 0
q[15]	= 0,020408
q[16]	= 0
q[17]	= 0
q[18]	= 0
q[19]	= 0
q[20]	= 0
q[21]	= 0,020408
q[22]	= 0
q[23]	= 0
q[24]	= 0
q[25]	= 0,020408
q[26]	= 0
q[27]	= 0,020408
q[28]	= 0
q[29]	= 0
q[30]	= 0
q[31]	= 0
q[32]	= 0,020408

$$- 1,4633 < H^{(20)} < 2,1888$$

Лабораторная работа №1

Произвольная часть текста:  
вы, верить, в, подлинное, существование, добра, и, зла, временами, люди, могут, ошибаться

Использованные буквы:  
., м, с, р, т, й, ц, у, к,

Порядок n-граммы:  
5 символов

Введенный символ: e  
Символ по счету: 10  
Номер эксперимента: 50

Неравенство для энтропии:  
 $1,46329361793241 < H < 2,18875843973145$

Двоичная таблица угаданных символов:

Вероятности:

q[1]	= 0,64
q[2]	= 0,04
q[3]	= 0,04
q[4]	= 0,02
q[5]	= 0,06
q[6]	= 0,04
q[7]	= 0,02
q[8]	= 0
q[9]	= 0,02
q[10]	= 0,02
q[11]	= 0
q[12]	= 0,02
q[13]	= 0
q[14]	= 0
q[15]	= 0
q[16]	= 0
q[17]	= 0
q[18]	= 0,02
q[19]	= 0,02
q[20]	= 0
q[21]	= 0
q[22]	= 0
q[23]	= 0
q[24]	= 0
q[25]	= 0
q[26]	= 0,04
q[27]	= 0
q[28]	= 0
q[29]	= 0
q[30]	= 0
q[31]	= 0
q[32]	= 0

$$- 0,8627 < H^{(30)} < 1,4803$$

Лабораторная работа №1

Произвольная часть текста:  
еским\_существам\_я\_надеюсь\_что

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: 51

Неравенство для энтропии:  
 $0,862658464396796 < H < 1,4803335295143$

Двоичная таблица угаданных символов:

10000000000000000000000000000000
10000000000000000000000000000000
00100000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:

q[1] = 0,76
q[2] = 0,06
q[3] = 0,04
q[4] = 0,04
q[5] = 0
q[6] = 0,02
q[7] = 0
q[8] = 0
q[9] = 0,02
q[10] = 0
q[11] = 0
q[12] = 0
q[13] = 0
q[14] = 0,02
q[15] = 0
q[16] = 0
q[17] = 0
q[18] = 0,02
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0,02
q[24] = 0
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0

Строка состояния:

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

- Надлишковість джерела відкритого тексту (мови) дорівнює:

$R = 1 - \frac{H_{\infty}}{H_0}$ , ( $H_0 = \log_2 32 = 5$ ) і характеризує величину можливого ущільнення тексту деякою схемою кодування символів без втрати його змісту.

$$1) 1,4368 < H^{(10)} < 2,2047$$

$$R = 1 - \frac{1,43680132713866}{5} \approx 0,71264$$

$$R = 1 - \frac{2,20471268847602}{5} \approx 0,55905$$

$$2) 1,4633 < H^{(20)} < 2,1888$$

$$R = 1 - \frac{1,46329361793241}{5} \approx 0,70734$$

$$R = 1 - \frac{2,18875843973145}{5} \approx 0,56225$$

$$3) 0,8627 < H^{(30)} < 1,4803$$

$$R = 1 - \frac{0,862658464396796}{5} \approx 0,82747$$

$$R = 1 - \frac{1,4803335295143}{5} \approx 0,70393$$

### **Висновок**

-Отже, виконуючи лабораторну роботу, я детально опрацювала теоретичний матеріал та звернула увагу на основні поняття. Написала програму для підрахунку частоти букв і частоти біграм в тексті. За допомогою програми CoolPinkProgram я отримала значення ентропії. Ці значення дали змогу проаналізувати, як змінюється ентропія при збільшенні розмірності аналізованих блоків. Таким чином, проведена робота дала можливість практично засвоїти поняття ентропії та надлишковості, зрозуміти їх застосування для аналізу текстової інформації.