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

КРИПТОГРАФІЯ

КОМП’ЮТЕРНИЙ ПРАКТИКУМ No1

Експериментальна оцінка ентропії на

символ джерела відкритого тексту.

Виконав:

Студент III курсу ФТІ

Групи ФБ-06

Сулима Олексій

Перевірила:

Селюх П.В.

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.

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 (10) H , (20) H , (30) H . 3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи:

	Key	Value
2	ч	0.01508685
3	а	0.0662817
4	с	0.04405879
5	т	0.05390995
6	ь	0.01913766
7		0.17461365
8	п	0.02284809
9	е	0.07247799
10	р	0.03473359
11	в	0.03850555
12	я	0.01776324
13	н	0.05414911
14	л	0.03827135
15	и	0.0539546
16	ю	0.00468394
17	з	0.01281236
18	ы	0.01374121
19	й	0.00830804
20	о	0.09555433
21	ж	0.00950284
22	к	0.02746951
23	м	0.02618441
24	д	0.02665082
25	у	0.02471373
26	ц	0.00229136
27	б	0.01447952
28	г	0.01404487
29	х	0.00708644
30	э	0.00294532
31	ф	0.00098144
32	щ	0.00249182
33	ъ	0.00020244
34	ё	6.3511E-05

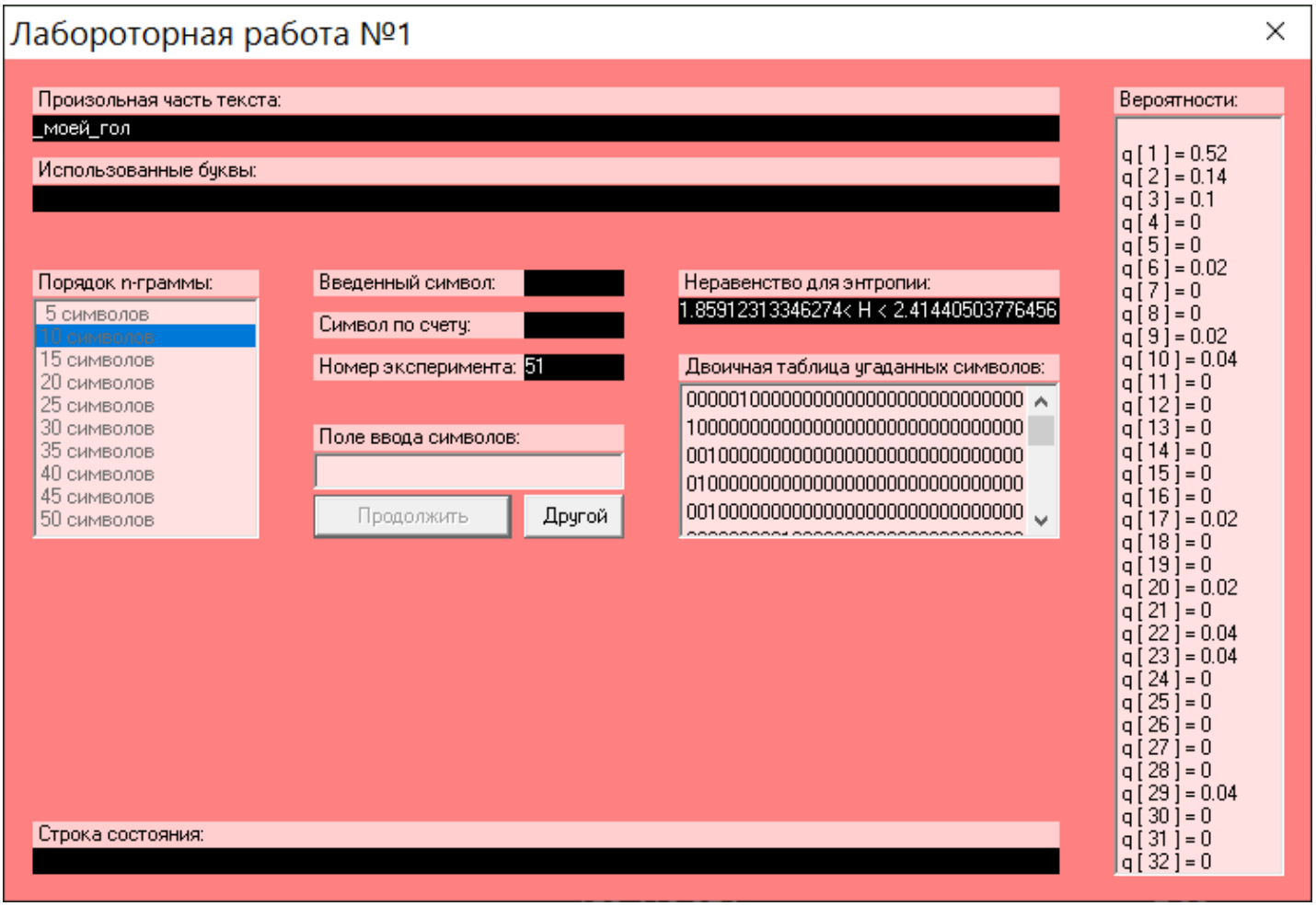
	A	B
1	Key	Value
2	ч	0.0182785
3	а	0.0803038
4	с	0.0533796
5	т	0.0653148
6	ь	0.0231863
7	п	0.0276817
8	е	0.087811
9	р	0.0420816
10	в	0.0466515
11	я	0.0215211
12	н	0.0656046
13	л	0.0463678
14	и	0.0653689
15	ю	0.0056748
16	з	0.0155229
17	ы	0.0166482
18	й	0.0100656
19	о	0.1157692
20	ж	0.0115132
21	к	0.0332808
22	м	0.0317238
23	д	0.0322889
24	у	0.029942
25	ц	0.0027761
26	б	0.0175427
27	г	0.0170161
28	х	0.0085856
29	э	0.0035684
30	ф	0.0011891
31	щ	0.003019
32	ъ	0.0002453
33	ё	7.695E-05

	Key	Value
2	('ч', 'а')	0.003047
3	('а', 'с')	0.007558
4	('с', 'т')	0.011703
5	('т', 'ь')	0.008
6	('ь', 'п')	0.001266
7	('п', 'е')	0.003434
8	('е', 'р')	0.007698
9	('р', 'в')	0.000505
10	('в', 'а')	0.007006
11	('а', 'я')	0.00268
12	('я', 'в')	0.001808
13	('в', 'н')	0.002502
14	('н', 'а')	0.012199
15	('а', 'ч')	0.001879
16	('а', 'л')	0.008967
17	('л', 'е')	0.004865
18	('е', 'и')	0.001927
19	('и', 'ю')	0.000369
20	('ю', 'л')	6.85E-05
21	('л', 'я')	0.001642
22	('в', 'ч')	0.000408
23	('ч', 'р')	0.00013
24	('р', 'е')	0.006367
25	('е', 'з')	0.002256
26	('з', 'в')	0.00119
27	('в', 'ы')	0.003265
28	('ы', 'ч')	0.000352
29	('а', 'й')	0.001074
30	('й', 'н')	0.000986
31	('н', 'о')	0.012012
32	('о', 'ж')	0.003114

	A	B
1	Key	Value
2	('ч', 'а')	0.00251
3	('а', 'с')	0.004658
4	('с', 'т')	0.00938
5	('т', 'ь')	0.006603
6	('ь', ' ')	0.012013
7	(' ', 'п')	0.015994
8	('п', 'е')	0.002833
9	('е', 'р')	0.005803
10	('р', 'в')	0.000362
11	('в', 'а')	0.005713
12	('а', 'я')	0.001918
13	('я', ' ')	0.011473
14	(' ', ' ')	0.006174
15	(' ', 'в')	0.016812
16	('в', ' ')	0.005627
17	(' ', 'н')	0.015964
18	('н', 'а')	0.010015
19	('а', 'ч')	0.000916
20	('а', 'л')	0.007125
21	('л', 'е')	0.003679
22	('е', ' ')	0.019033
23	(' ', 'и')	0.011727
24	('и', 'ю')	0.000301
25	('ю', 'л')	6.95E-06
26	('л', 'я')	0.001234
27	(' ', 'ч')	0.006191
28	('ч', 'р')	8.34E-05
29	('р', 'е')	0.005247
30	('е', 'з')	0.001184
31	('з', 'в')	0.000859
32	('в', 'ы')	0.002695

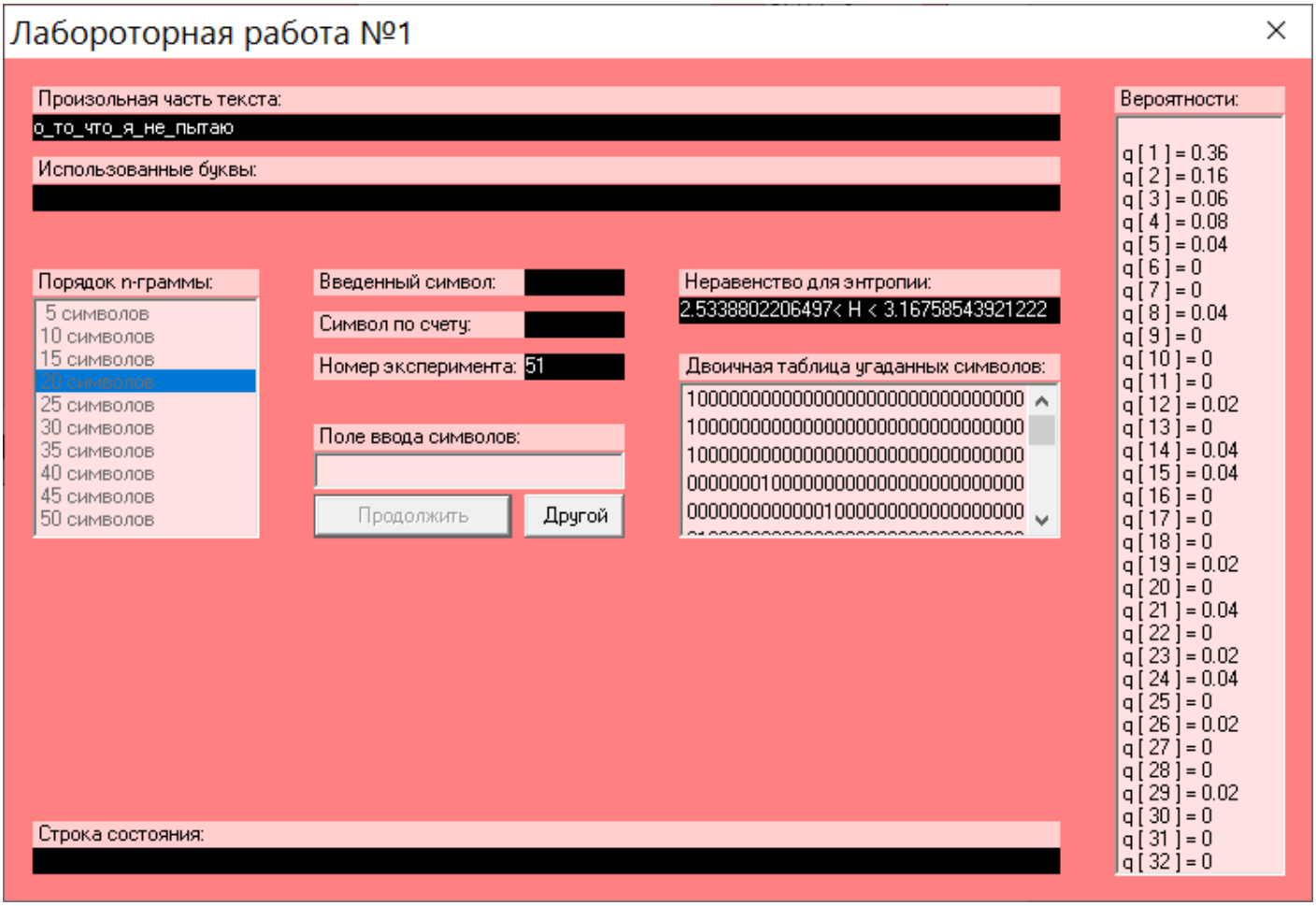
```
"C:\Program Files\python.exe" D:\cryptolabs\crypto-22-23\cp1\sulyma_fb06_cp1\lab1main.py
surplus H2 without spaces: 0.17913032778380056
surplus H1 without spaces: 0.11626901372807574
surplus H2 with spaces: 0.22139903712671205
surplus H1 with spaces: 0.14454577231594323
H1 with spaces: 4.315248275509783
H2 with spaces: 3.927570118444843
H1 without spaces: 4.4186549313596215
H2 without spaces: 4.104348361080997

Process finished with exit code 0
```

H10



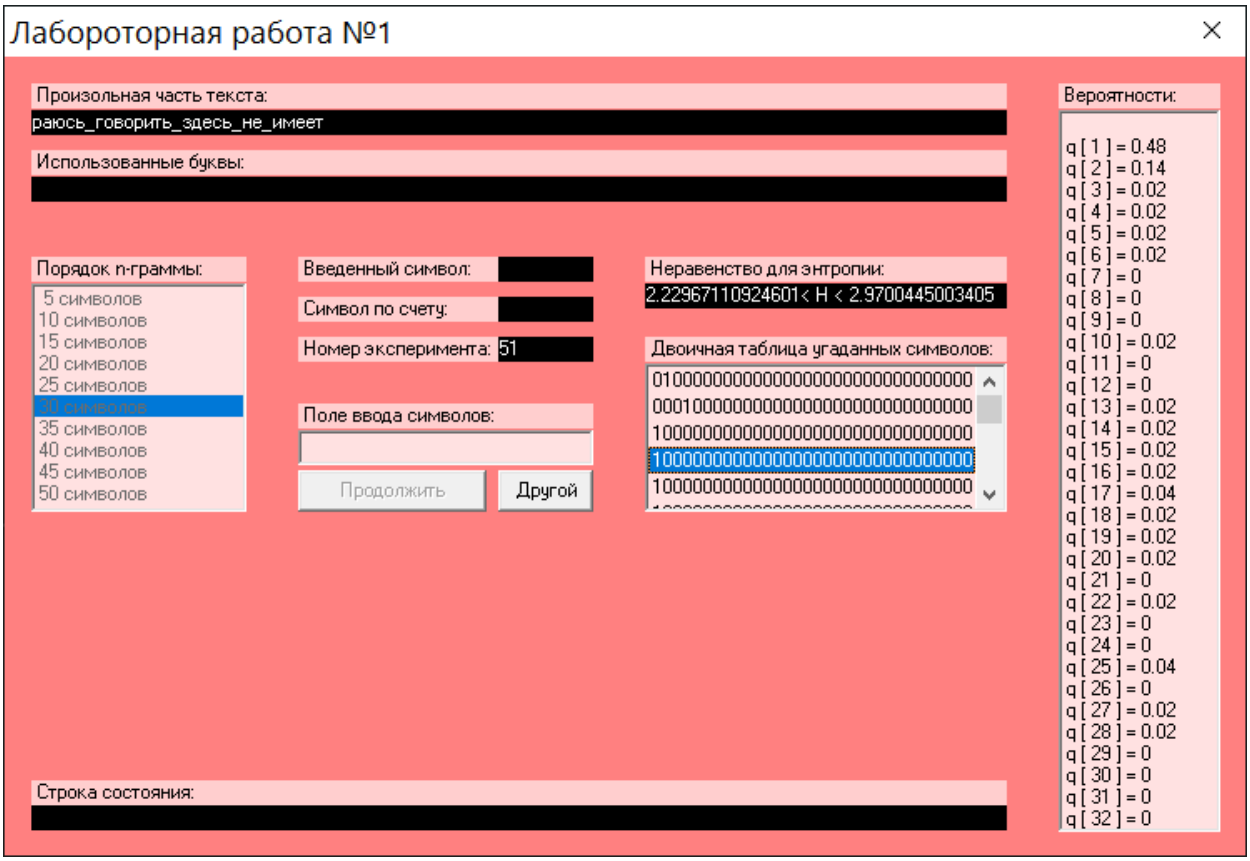
0.6912970018240869 < R10 < 0.5990937327380086

H20



0.5792551902063594 < R20 < 0.47402996656108864

H30



0.6297683786140637 < R30 < 0.5068311846422311

Висновки:
У ході виконання роботи, я навчився вимірювати частоти повторювання символів та біграм у тексті, визначати ентропію і надлишковість мови. Також помів, що у тексті без пробілів ентропія більша, ніж у тексті з пробілами. Тому перед шифруванням усі пробіли варто видаляти.