

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

Криптографія

Комп'ютерний практикум №1

**Експериментальна оцінка ентропії на символ джерела
відкритого тексту**

Виконав:

Студент групи ФБ-35

Ворона Сергій

Київ — 2025

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:

- 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.**

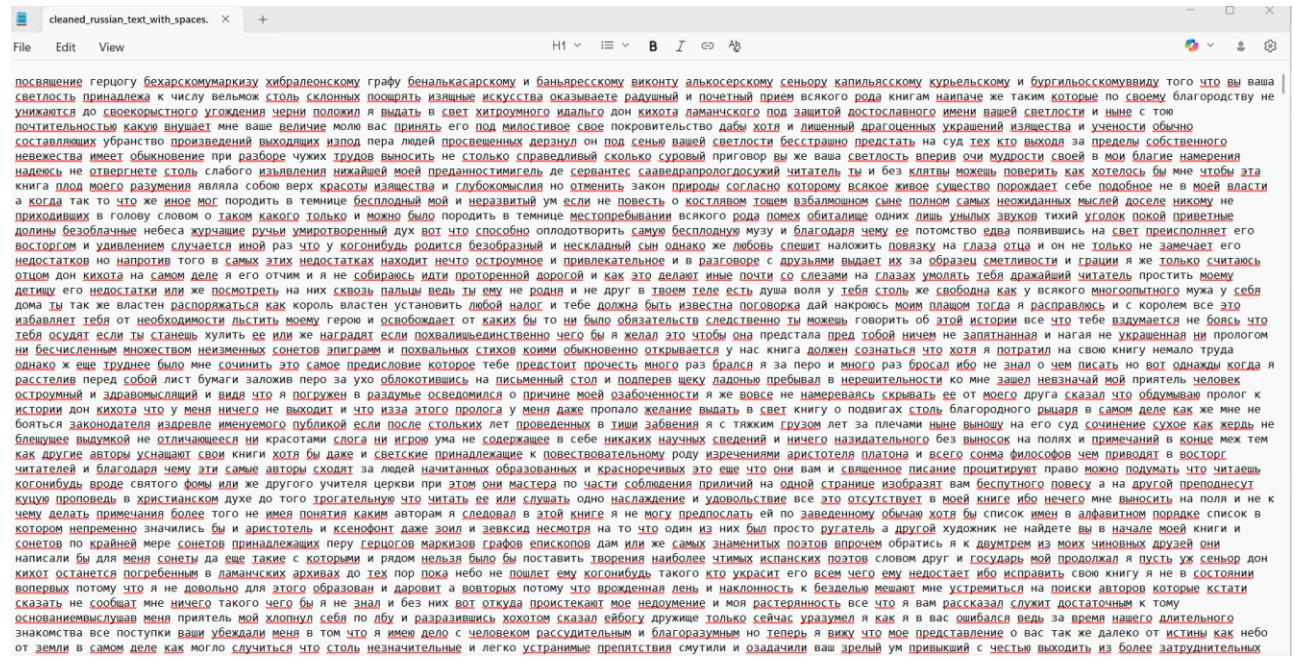
Текст щонайменше 1Мб:

russian_text.txt	07.10.2025 23:48	Text Document	1 808 KB
------------------	------------------	---------------	----------

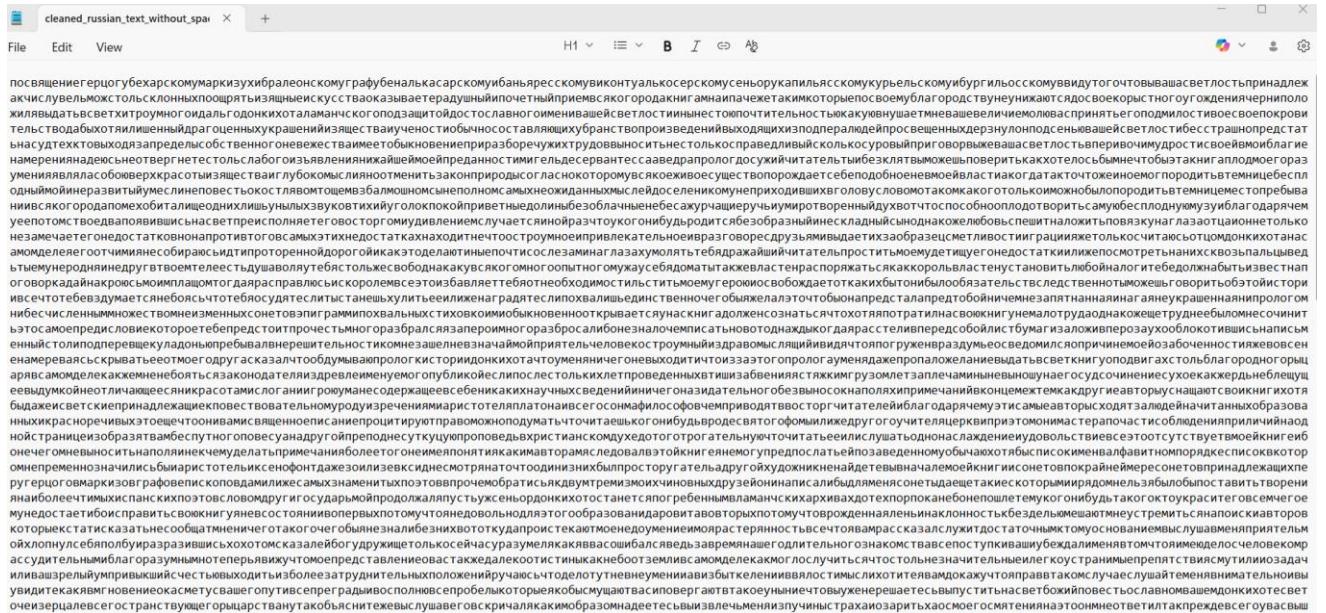
Текст очищений від знаків та цифр, окрім з пробілами та без:

cleaned_russian_text_with_spaces.txt	07.10.2025 23:49	Text Document	1 745 KB
cleaned_russian_text_without_spaces.txt	07.10.2025 23:50	Text Document	1 587 KB

З пробілами:



Без пробілів:



Підрахунок частот букв:

З пробілами:

Letter	Frequency
о	0,166603381
е	0,098869894
а	0,070682928
и	0,063710589
т	0,058735332
н	0,055288161
с	0,053786758
л	0,047107927
в	0,037728007
р	0,036087034
м	0,032989813
к	0,027008832
д	0,026765611
у	0,026427975
п	0,023190158
ь	0,020974485
я	0,017510894
ы	0,015225436
б	0,015173097
г	0,014239211
ч	0,0137579
з	0,01356907
ж	0,009028942
й	0,008262334
х	0,008221284
ш	0,006073344
ю	0,005925564
ц	0,003136218
щ	0,002988438
э	0,002712377
ф	0,000535702
ъ	0,000290429
ё	6,1575E-06

Без пробілів:

Letter	Frequency
о	0,118635
е	0,084813
а	0,076447
и	0,070477
т	0,066341
н	0,064539
с	0,056525
л	0,04527
в	0,043301
р	0,039585
м	0,032408
к	0,032116
д	0,031711
у	0,027826
п	0,025167
ь	0,021011
я	0,020862
ы	0,018269
б	0,018206
г	0,017086
ч	0,016508
з	0,016282
ж	0,010834
й	0,009914
х	0,009865
ш	0,007287
ю	0,00711
ц	0,003763
щ	0,003586
э	0,003255
ф	0,000643
ъ	0,000348
ё	7,39E-06

Підрахунок частот біграм (текст із пробілами):

Пари букв, що не перетинаються:

First_letter\Second_letter	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	
	0	0,00236	0,005811	0,014618	0,002662	0,008187	0,004955	0,002527	0,004366	0,012912	0	0,009743	0,002457	0,006868	0,015147	0,011496	0,014669	0,004142
а	0,014768	1,44E-05	0,000591	0,00314	0,000961	0,001757	0,001285	0,001043	0,003218	0,000213	0,00079	0,004883	0,006166	0,00361	0,004805	4,31E-05	0,000636	0,002321
б	0,000351	0,000484	2,05E-06	6,77E-05	0	2,45E-05	0,002241	2,05E-06	0,000686	0	1,85E-05	0,000979	4,93E-05	0,000480	0,002707	0	0,001418	
в	0,005347	0,005419	4,1E-06	3,49E-05	1,44E-05	0,000271	0,005045	0	0,000363	0,00272	0	9,65E-05	0,000768	0,00016	0,000792	0,006718	0,000277	0,000881
г	0,000642	0,001071	0	8,21E-06	0	0,000975	0,000172	0	0	0,000523	0	7,39E-05	0,000823	8,21E-06	0,000284	0,008495	0	0,000579
д	0,000702	0,004238	3,49E-05	0,000761	2,05E-06	4,1E-05	0,004259	4,1E-05	1,23E-05	0,002356	0	0,000131	0,000811	5,54E-05	0,001732	0,005638	0,000105	0,001158
е	0,019142	5,54E-05	0,001404	0,001328	0,003504	0,002738	0,001665	0,000907	0,001032	0,000156	0,001983	0,000792	0,005055	0,004491	0,007262	0,000287	0,000706	0,004105
ж	0,00046	0,001028	4,52E-05	0	6,16E-06	0,000969	0,004054	4,1E-06	0	0,001303	0	6,77E-05	1,23E-05	0	0,000803	4,31E-05	0	2,05E-06
з	0,001435	0,004928	0,00022	0,000835	0,000187	0,000601	0,000308	4,31E-05	9,03E-05	0,000376	0	0,000107	0,000287	0,00024	0,00148	0,000755	1,03E-05	0,000351
и	0,019649	0,00016	0,001186	0,001894	0,000515	0,001301	0,002108	0,000201	0,002036	0,000499	0,000932	0,002174	0,005277	0,002958	0,003173	0,000232	8,83E-05	0,000599
й	0,00635	0	2,87E-05	1,03E-05	6,16E-06	0,000216	1,23E-05	2,05E-06	6,16E-06	2,05E-05	0	9,44E-05	1,64E-05	3,28E-05	0,000482	1,03E-05	4,1E-06	2,05E-06
к	0,004454	0,006634	0	4,93E-05	4,1E-06	6,16E-06	0,0003	9,03E-05	2,05E-06	0,002999	0	8,21E-06	0,000772	0	0,000447	0,008187	4,1E-06	0,001252
л	0,005794	0,005022	1,23E-05	7,8E-05	0,000142	0,000119	0,00367	0,000394	0	0,006453	0	0,000199	0,000269	3,08E-05	0,000374	0,005111	1,23E-05	2,05E-06
м	0,007756	0,002022	5,95E-05	2,05E-06	4,72E-05	8,21E-06	0,003684	0	2,05E-06	0,002601	0	0,00014	0,000246	1,23E-05	0,00172	0,003955	3,69E-05	4,52E-05
н	0,004312	0,009333	2,05E-06	1,64E-05	1,64E-05	0,000474	0,009788	2,05E-05	8,21E-06	0,008163	0	0,000232	0	2,05E-06	0,002648	0,008153	8,21E-06	8,62E-05
о	0,024279	1,64E-05	0,004388	0,005597	0,004671	0,004692	0,002559	0,001669	0,001443	0,001203	0,002915	0,001582	0,005556	0,005018	0,006404	0,000234	0,01279	0,006063
п	3,69E-05	0,00134	0	0	0	0	0,001576	0	0	0,000819	0	6,57E-05	0,000647	0	0,000156	0,008633	4,1E-06	0,006519
р	0,000716	0,007069	7,18E-05	0,000289	0,00015	0,000445	0,005033	0,000314	7,39E-05	0,004557	0	0,000146	3,28E-05	5,13E-05	0,000827	0,006277	9,65E-05	4,1E-05

Пари букв, що перетинаються:

First_letter\Second_letter	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	
	0	0,00238	0,005832	0,014592	0,002675	0,008237	0,004843	0,002518	0,00439	0,012915	2,05E-06	0,009764	0,002364	0,006921	0,015051	0,011286	0,014782	0,004193
а	0,01478	1,13E-05	0,000595	0,003011	0,000966	0,001721	0,001384	0,001052	0,003245	0,000219	0,000774	0,004931	0,006207	0,003603	0,004808	4E-05	0,000598	0,002339
б	0,000322	0,000507	3,08E-06	7,39E-05	1,03E-06	2,16E-05	0,002164	2,05E-06	2,05E-06	0,000685	0	2,26E-05	0,000998	5,23E-05	0,000471	0,002715	0	0,001451
в	0,005308	0,005178	3,08E-06	3,18E-05	1,23E-05	0,000272	0,005128	0	0,000407	0,002779	0	8,83E-05	0,00079	0,000154	0,000835	0,006764	0,000271	0,000863
г	0,000664	0,001077	0	8,21E-06	0	0,000917	0,000172	0	1,03E-06	0,000556	0	7,77E-05	0,000839	7,18E-06	0,000284	0,008486	0	0,000604
д	0,000701	0,004286	3,59E-05	0,000754	3,08E-06	3,39E-05	0,004286	4,31E-05	1,13E-05	0,002356	1,03E-06	0,000142	0,000826	5,75E-05	0,0017	0,005683	0,000115	0,001113
е	0,019022	5,95E-05	0,00144	0,001329	0,00348	0,002785	0,001664	0,000902	0,00105	0,000144	0,001969	0,000837	0,005127	0,004464	0,007257	0,000296	0,00072	0,004246
ж	0,000421	0,001006	4,62E-05	0	7,18E-06	0,000992	0,004077	9,24E-06	0	0,001315	0	5,95E-05	1,13E-05	1,03E-06	0,000783	3,39E-05	0	4,11E-06
з	0,001398	0,004873	0,00021	0,000838	0,00018	0,000604	0,000281	5,13E-05	8E-05	0,000381	0	0,000117	0,000317	0,000242	0,001456	0,000704	7,18E-06	0,000365
и	0,019565	0,00022	0,001163	0,001894	0,000488	0,001288	0,000209	0,0002	0,002085	0,000541	0,000911	0,002172	0,005371	0,002848	0,00311	0,000225	8,42E-05	0,000565
й	0,006443	1,03E-06	1,23E-05	1,23E-05	5,13E-06	0,000227	7,18E-06	1,03E-06	6,16E-06	2,05E-05	0	9,13E-05	1,85E-05	2,87E-05	0,000461	1,44E-05	6,16E-06	1,03E-06
к	0,004501	0,006544	0	5,75E-05	4,11E-06	6,16E-06	0,000297	8,42E-05	1,03E-06	0,002975	0	4,11E-06	0,000755	0	0,000463	0,00812	3,08E-06	0,001301
л	0,005754	0,005019	1,23E-05	8,83E-05	0,000133	0,000125	0,003695	0,000412	4,11E-06	0,006465	0	0,000204	0,000265	3,28E-05	0,00038	0,005153	1,44E-05	2,05E-06
м	0,007963	0,001999	4,93E-05	1,03E-05	5,34E-05	9,24E-06	0,0003781	1,03E-06	3,08E-06	0,002695	0	0,000138	2,000222	1,13E-05	0,001702	0,003858	5,03E-05	4,11E-05
н	0,004333	0,00946	3,08E-06	1,44E-05	2,26E-05	0,000469	0,000905	1,85E-05	1,13E-05	0,000802	0	0,000218	0	4,11E-06	0,00264	0,008145	7,18E-06	6,67E-05
о	0,024104	1,64E-05	0,004464	0,005757	0,004694	0,004806	0,00267	0,001666	0,001411	0,00124	0,002889	0,001576	0,005488	0,004893	0,006442	0,000235	0,01293	0,005969
п	4,21E-05	0,001237	0	0	0	0	0,001547	0	0	0,000817	0	4,93E-05	0,000628	0	0,000149	0,008682	5,13E-06	0,00653
р	0,000714	0,007043	8,62E-05	0,000272	0,000141	0,000435	0,004989	0,000304	6,88E-05	0,004509	0	0,00016	3,8E-05	4,72E-05	0,00082	0,006285	8,31E-05	3,49E-05

Підрахунок частот біграм (текст без пробілів):

Пари букв, що не перетинаються:

First_letter\Second_letter	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	
	0,000318	0,001291	0,005142	0,001492	0,002948	0,00266	0,001492	0,004642	0,001689	0,000904	0,006921	0,007704	0,005216	0,007297	0,001153	0,002199	0,003362	
а	0,000623	2,46E-06	0,000113	2,46E-06	5,17E-05	0,002687	4,93E-06	9,85E-06	0,000837	0	5,42E-05	0,001273	6,16E-05	0,000611	0,003276	9,85E-06	0,001788	
в	0,006295	0,000148	0,000328	0,000217	0,000547	0,0006152	6,45E-05	0,000605	0,003783	0	0,000658	0,001044	0,000382	0,001399	0,008642	0,000842	0,001325	
г	0,00133	5,91E-05	4,43E-05	4,93E-06	0,000131	0,000227	2,46E-06	1,23E-05	0,000724	0	0,00014	0,001037	4,68E-05	0,000406	0,01025	5,17E-05	0,000739	
д	0,005105	6,65E-05	0,000956	2,46E-05	6,45E-05	0,0005204	4,43E-05	5,42E-05	0,002936	2,46E-06	0,000199	0,00101	0,000101	0,002175	0,006824	0,000195	0,001406	
е	0,																	

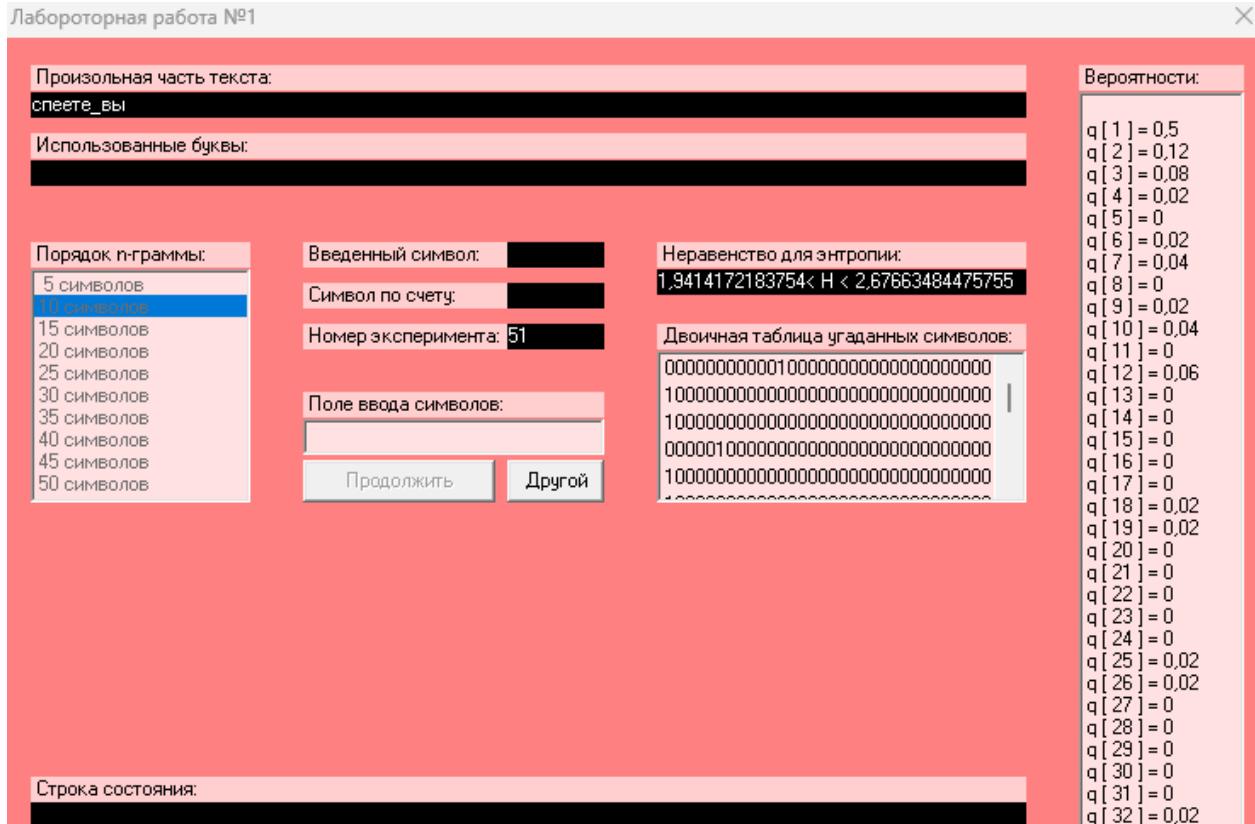
Пари букв, що перетинаються:

First_letter\Second_letter	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р															
а	0,000325	0,001319	0,005088	0,001485	0,002926	0,002193	0,001447	0,004416	0,001739	0,00093	0,007025	0,007722	0,005265	0,007416	0,001137	0,002242	0,003228															
б		0,000612	3,69E-06	0,000107	2,46E-06	4,06E-05	0,002616	3,69E-06	1,23E-05	0,000856	0	4,06E-05	0,001199	7,63E-05	0,000598	0,003315	9,85E-06	0,001749														
в			0,006312	0,000158	0,000352	0,000213	0,000578	0,006305	6,65E-05	0,0006	0,003675	0	0,000646	0,00105	0,000397	0,00143	0,008523	0,00084	0,001309													
г				0,001309	4,68E-05	6,53E-05	3,69E-06	0,001178	0,000232	2,46E-06	1,48E-05	0,000744	0	0,000123	0,001021	4,8E-05	0,000414	0,010245	6,9E-05	0,000739												
д					0,005166	6,77E-05	0,000968	1,72E-05	7,27E-05	0,005162	5,42E-05	5,54E-05	0,002906	1,23E-06	0,000212	0,001018	9,97E-05	0,002164	0,006855	0,000195	0,00135											
е						0,000324	0,002571	0,003638	0,004516	0,004468	0,002501	0,001449	0,001994	0,001853	0,002363	0,002044	0,006482	0,006461	0,010526	0,001877	0,003157	0,005811										
ж							0,00121	6,77E-05	8,99E-05	1,35E-05	0,001206	0,004911	1,11E-05	1,23E-05	0,001592	0	0,000101	2,09E-05	1,85E-05	0,001006	7,76E-05	3,57E-05	1,6E-05									
з								0,005873	0,00031	0,001148	0,000262	0,000823	0,000368	7,51E-05	0,000134	0,000505	0	0,000265	0,000417	0,000351	0,001939	0,000927	0,000150	0,0005								
и									0,000528	0,002373	0,00456	0,000896	0,002665	0,003267	0,000685	0,003129	0,000205	0,001093	0,003756	0,006738	0,004219	0,006023	0,001987	0,002473	0,001229							
й										0,000132	0,000268	0,000616	0,000207	0,000802	0,000132	0,00016	0,000116	0,000739	1,23E-06	0,000565	0,000127	0,000483	0,001081	0,000484	0,000688	0,000367						
к											0,007923	0,000232	0,000515	8,74E-05	0,000291	0,000493	0,000213	0,000123	0,003987	0	0,000227	0,000974	0,000215	0,001074	0,01064	0,000352	0,001685					
л												0,006135	0,000281	0,000624	0,000257	0,000728	0,004848	0,000526	0,000182	0,008247	0	0,000603	0,00039	0,000243	0,000951	0,006685	0,000452	9,6E-05				
м													0,002572	0,000367	0,000798	0,000246	0,00056	0,004711	0,000143	0,000224	0,004119	0	0,000701	0,000372	0,00034	0,002766	0,005318	0,000956	0,000305			
н													0,011409	0,00019	0,0004	9,6E-05	0,000687	0,011958	5,91E-05	0,000107	0,010051	0	0,001582	7,02E-05	0,000161	0,003553	0,01026	0,000493	0,00169			
о														0,000291	0,006519	0,009649	0,006119	0,007111	0,004152	0,002596	0,002512	0,003328	0,003466	0,00324	0,007017	0,007076	0,010148	0,002549	0,004048	0,007853		
п															0,001488	1,23E-06	2,46E-06	1,23E-06	0	0,001857	0	2,46E-06	0,000988	0	0,03E-05	0,000756	2,46E-06	0,000182	0,010423	1,11E-05	0,007835	
р																0,008488	0,000121	0,000393	0,000175	0,000586	0,0006008	0,000368	9,97E-05	0,005493	0	0,000254	6,77E-05	8,13E-05	0,001063	0,007587	0,000196	6,53E-05

Підрахунок ентропії та надлишковості:

	Ентропія	Надлишковість
H1 R1 (3 пробілами)	4,356922	0,143596344
H1 R1 (Без пробілами)	4,448118	0,118205739
H2 R2 (Перетинаються, з пробілами)	3,947526	0,224067886
H2 R2 (Перетинаються, без пробілів)	4,127006	0,181862832
H2 R2 (Не перетинаються, з пробілами)	3,947499	0,224073088
H2 R2 (Не перетинаються, без пробілів)	4,127232	0,181818161

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



Лабораторная работа №1

Произвольная часть текста:	человек_имеет_свой_з		
Использованные буквы:			
Порядок n-грамм:	Введенный символ:	Неравенство для энтропии: 1.39211889621353 < H < 2.0592247882468	
5 символов			
10 символов			
15 символов			
20 символов			
25 символов			
30 символов			
35 символов			
40 символов			
45 символов			
50 символов			
	Символ по счету:		
	Номер эксперимента: 51		
	Поле ввода символов:	Двоичная таблица угаданных символов:	
		01000000000000000000000000000000 10000000000000000000000000000000 00000000010000000000000000000000 10000000000000000000000000000000 10000000000000000000000000000000	
		~~~~~ ~~~~~	
	Продолжить	Другой	
Строка состояния:			

Вероятности:

```

q[1] = 0.62
q[2] = 0.14
q[3] = 0.02
q[4] = 0.02
q[5] = 0.02
q[6] = 0
q[7] = 0.04
q[8] = 0
q[9] = 0.02
q[10] = 0.04
q[11] = 0
q[12] = 0
q[13] = 0
q[14] = 0
q[15] = 0
q[16] = 0
q[17] = 0.02
q[18] = 0
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0.02
q[23] = 0
q[24] = 0
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0.04
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0

```

Лабораторная работа №1

Произвольная часть текста:	не_стали_бы_связывать_себя_сл		
Использованные буквы:			
Порядок n-грамм:	Введенный символ:	Неравенство для энтропии: 1.57702940353673 < H < 2.26622721260073	
5 символов			
10 символов			
15 символов			
<b>20 символов</b>			
25 символов			
30 символов			
35 символов			
40 символов			
45 символов			
50 символов			
	Символ по счету:		
	Номер эксперимента: 51		
	Поле ввода символов:	Двоичная таблица угаданных символов:	
		10000000000000000000000000000000 10000000000000000000000000000000   10000000000000000000000000000000 0000001000000000000000000000000 10000000000000000000000000000000	
		~~~~~ ~~~~~	
	Продолжить	Другой	
Строка состояния:			

Вероятности:

```

q[1] = 0.58
q[2] = 0.16
q[3] = 0
q[4] = 0.02
q[5] = 0.04
q[6] = 0.02
q[7] = 0.02
q[8] = 0.02
q[9] = 0
q[10] = 0
q[11] = 0
q[12] = 0
q[13] = 0
q[14] = 0.02
q[15] = 0.02
q[16] = 0.02
q[17] = 0.02
q[18] = 0
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0.02
q[23] = 0
q[24] = 0
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0
q[29] = 0
q[30] = 0
q[31] = 0.04
q[32] = 0

```

1.9414172183754 < H⁽¹⁰⁾ < 2.67663484475755

0.46467303104849 < R⁽¹⁰⁾ < 0.611716556325

1.39211889621353 < H⁽²⁰⁾ < 2.0592247882468

0.58815504235064 < R⁽²⁰⁾ < 0.721576220757294

1.57702940353673 < H⁽³⁰⁾ < 2.26622721260073

0.546754557479854 < R⁽³⁰⁾ < 0.684594119292654

Висновок:

У цій лабораторній роботі ми досліджували різні моделі джерела. Впродовж виконання я розібрався із поняттям ентропії та надлишковості. Навчився розраховувати частоти букв та біграм та на її основі підраховувати значення ентропії та надлишковості мови. Отримані результати дають можливість побачити статистичні властивості мови та її надлишковість.