

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

*КРИПТОГРАФІЯ*

*Комп'ютерний практикум*

*Робота № 1*

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

*Виконав:*

*студент гр. ФБ-02*

*Шубін Д.Ю*

**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  $H_1$  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.*
- 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:**

- 1) Частоти букв у тексті з пробілами

а	0,06953
б	0,01352
в	0,03372
г	0,0184
д	0,02625
е	0,06564
ё	0,00001
э	0,00232
ж	0,00813
з	0,01596
и	0,05901
ы	0,01364
й	0,00757
к	0,02668
л	0,04597
м	0,02668
н	0,05415
о	0,09251
п	0,02464
р	0,0472
с	0,04492
т	0,04783
у	0,02546
ф	0,00228
х	0,00696
ц	0,00236
ч	0,01241
ш	0,00718
щ	0,00248
ъ	0,00024
ь	0,01626
ю	0,00389
я	0,01658
	0,15963

2) Частоти букв у тексті без пробілів

	а	0,08274
	б	0,01608
	в	0,04013
	г	0,02189
	д	0,03123
	е	0,07811
	ё	0,00002
	э	0,00276
0	ж	0,00967
1	з	0,019
2	и	0,07022
3	ы	0,01623
4	й	0,009
5	к	0,03175
5	л	0,0547
7	м	0,03174
3	н	0,06443
9	о	0,11009
0	п	0,02932
1	р	0,05617
2	с	0,05345
3	т	0,05691
4	у	0,0303
5	ф	0,00271
5	х	0,00828
7	ц	0,00281
3	ч	0,01477
9	ш	0,00855
0	щ	0,00295
1	ъ	0,00029
2	ь	0,01934
3	ю	0,00463
4	я	0,01973

3) Частота біграм у тексті з пробілами

[illegible]

4) Частота біграм в тексті з пробілами, з 2-им кроком

[illegible]

5) Частота біграм в тексті без пробілів

[illegible]

6) Частота біграм в тексті без пробілів, з 2-им кроком

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	
1	#																																		
2	#	0.00006	0.00159	0.00091	0.00234	0.00336	0.00191	0.00001	0.00001	0.00001	0.00012	0.00496	0.00115	0.0	0.00071	0.0058	0.01084	0.00463	0.00555	0.01682	0.00354	0.00684	0.00044	0.00585	0.00079	0.00026	0.00117	0.00017	0.00154	0.00039	0.00003	0	0	0.00096	0.00238
3	#	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	
4	#	0.00059	0.00029	0.00049	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	0.00009	
5	#	0.00021	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	
6	#	0.00023	0.00021	0.00033	0.00037	0.00049	0.00162	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	
7	#	0.00023	0.00021	0.00033	0.00037	0.00049	0.00162	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	
8	#	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	#	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
10	#	0.0145	0.00004	0.0001	0.00006	0.00019	0.00019	0.00001	0.00001	0.00001	0.00001	0.00001	0.0016	0.0	0.00019	0.00003	0.00004	0.00009	0.00017	0.00015	0.00002	0.00008	0.00026	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	
11	#	0.00745	0.00022	0.00027	0.0006	0.00106	0.0006	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00003	0.00006	0.00015	0.00004	0.00014	0.00019	0.00023	0.00049	0.00018	0.00008	0.00066	0	0.00003	0.00003	0	0.00003	0.00016	0	0.00008	0.00028		
12	#	0.00129	0.00012	0.00012	0.																														
13	#	0.00007	0.00043	0.00143	0.00001	0.00017	0.00012	0.00001	0.00001	0.00001	0.00014	0.0029	0.00016	0.	0.00147	0.00041	0.01182	0.00129	0.00036	0.00038	0.00042	0.00119	0.00093	0.00021	0.00099	0.00099	0.00091	0.00032	0.00008	0.00001	0.	0.	0.00008		
14	#	0.00012	0.00002	0.00008	0.0.0	0.0005	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	
15	#	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	0.00019	
16	#	0.00083	0.00024	0.0008	0.00113	0.00079	0.00495	0.00001	0.00014	0.00036	0.00067	0.00007	0.001	0.	0.00096	0.00101	0.00034	0.00121	0.00076	0.00091	0.00052	0.00033	0.00034	0.00182	0.00041	0.00001	0.00001	0.00078	0.0002	0.	0.00032	0.00088	0.00181		
17	#	0.00351	0.00045	0.00071	0.00045	0.00041	0.00416	0.	0.00009	0.00006	0.00034	0.00528	0.00114	0.00001	0.00071	0.0003	0.00035	0.0012	0.00427	0.00102	0.00038	0.00112	0.00031	0.00324	0.00005	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008	0.00008		
18	#	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013	0.00013		
19	#	0.00019	0.00495	0.00099	0.00056	0.00513	0.00224	0.	0.00044	0.00213	0.00033	0.00171	0.001	0.00359	0.0049	0.00775	0.00069	0.01205	0.0026	0.00048	0.00012	0.00029	0.000929	0.00068	0.00008	0.00107	0.00013	0.00052	0.00013	0.00022	0.	0.	0.00015		
20	#	0.00222	0.																																
21	#	0.00012	0.00002	0.00002	0.	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001		
22	#	0.00016	0.00018	0.00147	0.00003	0.00056	0.00066	0.00001	0.00001	0.00009	0.00014	0.00297	0.00003	0.	0.00041	0.00043	0.00144	0.00105	0.00038	0.00281	0.00044	0.00075	0.01109	0.00108	0.00011	0.00025	0.0006	0.00052	0.00011	0.	0.00003	0.00016	0.00017		
23	#	0.00055	0.00024	0.00062	0.00026	0.00038	0.00014	0.00001	0.00001	0.00001	0.00002	0.00049	0.00019	0.00001	0.00103	0.00046	0.00026	0.00108	0.01373	0.00075	0.00038	0.00185	0.00072	0.00174	0.00004	0.00006	0.00006	0.00006	0.00006	0.00006	0.00006	0.00006	0.00006		
24	#	0.00081	0.00001	0.00014	0.0014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014		
25	#	0.00012	0.	0.00001	0.	0.	0.00003	0.	0.	0.	0.00004	0.00004	0.	0.	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001		
26	#	0.00044	0.00008	0.00035	0.0001	0.00014	0.00073	0.00001	0.00001	0.00002	0.00012	0.00033	0.00001	0.00001	0.00019	0.00029	0.00034	0.00004	0.00031	0.00034	0.00027	0.00003	0.00003	0.00014	0.00007	0.00002	0.00002	0.00001	0.00006	0.00004	0.00004	0.	0.	0.00003	
27	#	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001		
28	#	0.00022	0.00002	0.00002	0.	0.00001	0.00002	0.	0.	0.00018	0.	0.	0.	0.	0.00074	0.00005	0.	0.00007	0.00004	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003		
29	#	0.0011	0.	0.00005	0.	0.	0.00113	0.	0.	0.	0.00112	0.	0.	0.	0.00076	0.00078	0.00005	0.00004	0.00005	0.00002	0.00008	0.00005	0.00007	0.00025	0.	0.	0.	0.	0.	0.	0.	0.	0.00003		
30	#	0.0004	0.	0.00004	0.	0.	0.00041	0.	0.	0.	0.00038	0.	0.	0.	0.00008	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.00002	
31	#	0.	0.	0.	0.	0.	0.00007	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.00002	
32	#	0.00015	0.00001	0.00001	0.00005	0.00016	0.00009	0.	0.00008	0.00006	0.00004	0.00072	0.	0.	0.00148	0.00019	0.00001	0.00001	0.00007	0.00003	0.00012	0.00003	0.00009	0.00032	0.00018	0.00004	0.00015	0.00009	0.00008	0.00001	0.	0.	0.00048		
33	#	0.00001	0.00004	0.00021	0.0001	0.00001	0.00001	0.	0.	0.00001	0.00008	0.00001	0.00001	0.	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	
34	#	0.00015	0.00001	0.00021	0.0001	0.00001	0.00001	0.	0.	0.00001	0.00023	0.00071	0.00007	0.00001	0.00005	0.00019	0.00013	0.00008	0.00026	0.00017	0.00024	0.00019	0.000213	0.00004	0.00008	0.00027	0.00008	0.00046	0.00001	0.00019	0.	0.	0.0001	0.00023	







**Висновок:** у ході лабораторної роботи, я на практиці розібрався з поняттям ентропії, надлишковості, реалізувавши їх на мові програмування Python. Ознайомився з вивченням та порівнянням різних моделей джерела відкритого тексту для наближеного визначення ентропії, набув практичних навичок щодо оцінки ентропії на символу джерела.