

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**  
**“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ”**  
**ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ**

**Криптографія**

**КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1**  
**«Експериментальна оцінка ентропії на символ джерела**  
**відкритого тексту»**

**ФБ-32 Дорошенко Ілля**

**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.

### 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:

Для дослідження було обрано текст російською мовою «ТЕХТ.txt». Згідно з методичними вказівками, текст пройшов попередню фільтрацію:

- Усі символи, крім літер, були вилучені або замінені на пробіли.
- Прописні літери замінені на відповідні рядкові.
- Послідовності пробілів трактуються як один пробіл.
- Буква «ё» замінена на «е», а «ъ» на «ь».
- Алфавіт дослідження склав 32 літери (без пробілу) або 33 символи (з пробілом).

При підрахунку біграм було реалізовано два підходи:

1. **Крок 1:** пари букв, що перетинаються (більш точна статистика).
2. **Крок 2:** пари букв, що не перетинаються.

На основі роботи програми було отримано такі значення ентропії та надлишковості:

А	В	С
Параметр	З пробілами	Без пробілів
$H_1$	4,383006841	4,468568723
$H_2$ (step 1)	3,975230577	4,150663708
$H_2$ (step 2)	3,974254627	4,149794329
$R(H_1)$	0,123398632	0,098023485
$R(H_2 \text{ step } 1)$	0,204953885	0,162192322

Таблиці частот символів:

(З пробілом)

	А	В
1	Символ	Частота
2		0,162346
3	о	0,095222
4	а	0,070264
5	е	0,066722
6	и	0,055668
7	н	0,054587
8	т	0,047573
9	с	0,043704
10	л	0,042398
11	в	0,03856
12	р	0,038175
13	к	0,030044
14	д	0,025471
15	м	0,024757
16	у	0,024013
17	п	0,021519
18	я	0,019391
19	г	0,017368
20	ь	0,016753
21	ы	0,015897
22	з	0,014918
23	б	0,014467
24	ч	0,011416
25	й	0,009648
26	ж	0,008485
27	ш	0,00791
28	х	0,007145
29	ю	0,005433
30	ц	0,003388
31	э	0,002533
32	щ	0,002349
33	ф	0,001875

(Без пробілу)

	А	В
1	Символ	Частота
2	о	0,113677
3	а	0,083882
4	е	0,079653
5	и	0,066457
6	н	0,065167
7	т	0,056793
8	с	0,052174
9	л	0,050615
10	в	0,046034
11	р	0,045574
12	к	0,035867
13	д	0,030408
14	м	0,029556
15	у	0,028667
16	п	0,025689
17	я	0,02315
18	г	0,020735
19	ь	0,02
20	ы	0,018979
21	з	0,017809
22	б	0,017271
23	ч	0,013629
24	й	0,011518
25	ж	0,01013
26	ш	0,009443
27	х	0,00853
28	ю	0,006486
29	ц	0,004045
30	э	0,003024
31	щ	0,002805
32	ф	0,002239

Частота біграм з перекриттям з пробілами:

	А	В	С	Д	Е	Г	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ
1	А	В	С	Д	Е	Г	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ
2	В	С	Д	Е	Г	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ	
3	С	Д	Е	Г	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ		
4	Д	Е	Г	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ			
5	Е	Г	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ				
6	Г	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ					
7	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ						
8	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ							
9	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ								
10	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ									
11	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ										
12	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ											
13	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ												
14	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ													
15	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ														
16	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ															
17	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ																
18	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ																	
19	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ																		
20	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ																			
21	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ																				
22	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ																					
23	Ц	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ																						
24	Ч	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ																							
25	Ш	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ																								
26	Щ	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ																									
27	Ъ	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ																										
28	Ы	Э	Ю	Я	АВ	АС	АД	АЕ	АГ																											
29	Э	Ю	Я	АВ	АС	АД	АЕ	АГ																												
30	Ю	Я	АВ	АС	АД	АЕ	АГ																													
31	Я	АВ	АС	АД	АЕ	АГ																														
32	АВ	АС	АД	АЕ	АГ																															
33	АС	АД	АЕ	АГ																																
34	АД	АЕ	АГ																																	
35	АЕ	АГ																																		
36	АГ																																			

Без пробілів:

1	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	
2	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	
2	a	0.000388	0.001583	0.005821	0.001719	0.003269	0.001533	0.001592	0.001572	0.000906	0.006516	0.010889	0.004993	0.007899	0.001472	0.002712	0.003733	0.00633	0.006404	0.005553	0.000713	0.002177	0.000158	0.001298	0.001275	0.00028	0	0	0.000347	0.000928	0.003512		
3	b	0.001368	5.53E-05	5.94E-05	2.23E-05	2.04E-05	0.002359	9.28E-06	7.71E-06	0.000995	0	0.000303	0.000898	8.17E-05	0.000316	0.002703	5.57E-06	0.001246	8.65E-05	1.11E-05	0.001383	0	5.75E-05	7.43E-06	3.34E-05	7.43E-06	0.000198	0.004217	0.000339	3.34E-05	7.43E-06	0.000359	
4	c	0.006987	0.000249	0.000464	0.000412	0.000707	0.005125	2.97E-05	0.000759	0.004199	0	0.000733	0.001296	5.000349	0.002116	0.008765	0.001062	0.000802	0.004247	0.000796	0.001088	1.67E-06	7.98E-05	7.24E-05	0.000239	0.001548	9.28E-06	0.003154	0.000212	0.00034	5.57E-06	0.000173	
5	d	0.00124	4.27E-05	0.000165	1.3E-05	0.001099	0.000698	0	7.61E-05	0.001034	0	0.000187	0.001927	5.01E-05	0.000444	0.010765	9.84E-05	0.001628	0.000156	4.83E-05	0.000962	7.43E-06	3.71E-06	0	3.34E-05	1.3E-05	0	3.71E-06	0	2.04E-05	1.11E-05	9.28E-06	
6	e	0.004798	9.47E-05	0.001175	9.28E-05	5.57E-05	0.005511	2.23E-05	5.2E-05	0.003328	0	0.000336	0.000743	0.000149	0.002072	0.004217	0.000238	0.00246	0.00046	0.000388	0.001938	3.34E-05	6.5E-05	0.000195	8.17E-05	0.000132	0	0.000078	0.000913	2.04E-05	1.86E-05	0.00044	
7	f	0.000195	0.002196	0.003445	0.004544	0.003724	0.002047	0.001337	0.001964	0.001448	0.002721	0.002189	0.007455	0.002534	0.009441	0.001611	0.001139	0.000339	0.000548	0.002523	0.000583	6.87E-05	0.003417	0.000334	0.002561	0.000963	0.000772	0	0	0.000219	0.000414	0.000611	
8	g	0.001593	6.68E-05	2.97E-05	3.34E-05	0.000756	0.004418	1.49E-05	9.28E-06	0.001314	0	0.000113	1.86E-06	2.97E-05	0.001244	5.75E-05	1.49E-05	0	3.16E-05	1.3E-05	0.000228	0	5.57E-06	0	5.94E-05	0	0	0	4.08E-05	1.86E-05	2.41E-05	1.3E-05	
9	h	0.006462	0.0002	0.000978	0.000562	0.000869	0.000258	0.000208	0.000115	0.000423	1.86E-06	0.000338	0.000241	0.000347	0.001906	0.000904	0.000191	0.000238	0.000342	7.24E-05	0.00033	3.71E-06	5.57E-06	1.3E-05	4.64E-05	2.97E-05	1.86E-06	0.000631	0.001311	3.34E-05	0.000134	0.000613	
10	i	0.000388	0.001426	0.00525	0.001285	0.002979	0.003271	0.000613	0.002545	0.001993	0.001689	0.003195	0.006128	0.003937	0.005701	0.002072	0.002562	0.001437	0.003532	0.004713	0.000553	0.000145	0.002138	0.00173	0.00197	0.000587	0.000137	0	3.71E-06	0.000256	0.000349	0.002055	
11	j	0.0002	0.000414	0.000739	0.000373	0.000718	0.000126	0.000123	0.000161	0.000601	1.86E-06	0.000802	0.000416	0.000448	0.000967	0.000436	0.000848	0.000416	0.000137	0.000176	0.00065	0.000247	5.94E-05	5.01E-05	5.94E-05	0.000349	0.000189	1.11E-05	0	0	6.87E-05	1.11E-05	9.28E-05
12	k	0.009023	0.000646	0.000622	0.00021	0.000239	0.000648	0.000187	9.47E-05	0.003297	1.86E-06	0.000449	0.000487	0.000203	0.003219	0.010191	0.000394	0.002135	0.000624	0.000726	0.001914	2.66E-05	5.01E-05	3.16E-05	0.000156	2.78E-05	3.71E-06	0	0	9.1E-05	2.23E-05	7.43E-05	
13	l	0.00775	0.000416	0.000807	0.000466	0.000807	0.004821	0.000431	0.000191	0.007958	1.86E-06	0.001858	0.000264	0.000128	0.001104	0.008537	0.000924	0.000358	0.0002374	0.000369	0.001544	3.53E-05	2.97E-05	5.57E-06	0.000668	7.43E-05	5.57E-06	0.001442	0.004095	9.65E-05	0.000978	0.000273	
14	m	0.00351	0.000295	0.000817	0.000446	0.00036	0.003349	0.000121	0.000199	0.0048	0	0.000709	0.00029	0.000273	0.002029	0.004778	0.00103	0.000252	0.001073	0.00033	0.002794	5.2E-05	9.28E-05	4.64E-05	0.00047	8.17E-05	5.57E-06	0.000737	2.78E-05	7.98E-05	7.61E-05	0.000592	
15	n	0.01224	0.000345	0.000553	0.000343	0.001533	0.010356	5.94E-05	0.000212	0.008182	0	0.00052	5.9E-05	0.000124	0.003989	0.010286	0.000681	0.000149	0.000999	0.000223	0.001326	2.41E-05	5.57E-05	0.000679	0.000317	6.5E-05	0.000173	0.000809	0.001184	3.53E-05	0.00018	0.004002	
16	o	0.00023	0.000546	0.012032	0.006155	0.00628	0.003113	0.002304	0.001658	0.002294	0.004426	0.003223	0.008372	0.006562	0.01043	0.002344	0.003542	0.00788	0.009907	0.009029	0.000598	0.000692	0.000794	0.000149	0.000299	0.00165	0.000158	0	1.86E-06	0.000408	0.000882	0.001088	
17	p	0.001374	1.86E-06	1.86E-06	1.67E-05	1.86E-06	0.002887	0	0	0.000891	0	7.98E-05	0.000845	0	6.31E-05	0.010124	9.47E-05	0.007046	1.11E-05	6.13E-05	0.000809	1.86E-06	0	1.86E-06	9.28E-06	1.49E-05	0	0.000232	0.000696	3.71E-06	0	0.000421	
18	q	0.009937	0.000208	0.000436	0.000368	0.00052	0.007076	0.000438	6.68E-05	0.006284	1.86E-06	0.000498	8.91E-05	0.000306	0.000898	0.008233	0.000243	9.28E-05	0.000477	0.000896	0.003688	4.83E-05	0.000174	0.000115	0.000282	7.61E-05	0.001698	0.00095	1.86E-05	0.000247	0.001272		
19	r	0.001736	0.000213	0.002274	0.000134	0.000512	0.004204	0.000106	0.000124	0.002064	1.86E-06	0.003777	0.000289	0.001247	0.001227	0.003087	0.002307	0.000446	0.001273	0.01269	0.00103	5.2E-05	0.000213	2.78E-05	0.000429	9.28E-05	1.86E-06	0.000471	0.00093	6.87E-05	0.000152	0.000627	
20	s	0.006491	0.000321	0.002992	0.00018	0.000407	0.005591	6.5E-05	0.000139	0.004329	3.71E-06	0.000744	0.000395	0.000243	0.00155	0.016072	0.000546	0.003432	0.00145	0.000379	0.002222	4.64E-05	5.94E-05	0.000143	0.000401	2.6E-05	2.6E-05	0.001832	0.005853	0.000145	2.24E-05	0.000655	
21	t	0.000197	0.000839	0.001598	0.001695	0.002599	0.000282	0.001561	0.001175	0.000668	0.000111	0.001858	0.001867	0.001522	0.001181	0.000447	0.001312	0.000843	0.002354	0.002018	0.000167	4.83E-05	0.00047	1.3E-05	0.001203	0.000936	0.000236	0	0	0.000123	0.001188	0.000156	
22	u	0.000217	3.71E-06	2.6E-05	5.57E-06	5.57E-06	0.000165	0	7.43E-06	0.000772	0	2.78E-05	0.000115	1.11E-05	2.41E-05	7.8E-05	2.23E-05	0.000579	2.97E-05	1.67E-05	9.28E-05	5.57E-06	0	0	1.86E-06	0	0	2.23E-05	1.86E-06	0	5.57E-06		
23	v	0.001194	0.000195	0.000484	0.000175	0.000152	0.000106	4.27E-05	5.98E-05	0.000483	1.86E-06	0.000251	0.000217	0.000208	0.000408	0.002931	0.000292	0.000219	0.000438	0.000145	0.000221	2.41E-05	9.28E-06	7.43E-06	7.8E-05	2.6E-05	1.86E-06	0	3.71E-06	4.08E-05	0	2.78E-05	
24	w	0.000674	9.28E-06	5.2E-05	3.9E-05	2.04E-05	0.001288	1.86E-06	9.28E-06	0.000288	0	0.000219	9.28E-06	2.23E-05	5.57E-05	0.000516	3.9E-05	5.57E-06	4.08E-05	1.86E-05	0.000583	0	3.71E-06	0	0	0	0.000139	0	3.71E-06	0	3.71E-06		
25	x	0.002525	1.49E-05	1.11E-05	5.57E-06	5.57E-06	0.003503	1.86E-06	5.57E-06	0.001401	0	0.000277	4.27E-05	5.57E-06	0.000555	0.000134	1.3E-05	3.34E-05	2.78E-05	0.00406	0.000679	0	1.86E-06	0	3.71E-06	8.54E-05	0	0	0.00023	3.71E-06	0	3.71E-06	
26	y	0.00163	1.67E-05	3.16E-05	0	1.3E-05	0.002569	0	1.86E-06	0.002283	0	0.00049	0.000542	3.34E-05	0.000408	0.000395	1.87E-05	5.57E-06	1.3E-05	0.000191	0.00034	1.86E-06	0	1.86E-06	5.57E-06	0	0	0.000382	0	0	0		
27	z	0.000483	0	0	0	1.86E-06	0.001403	0	0	0.000772	0	0	0	0	3.53E-05	0	0	3.71E-06	1.86E-06	1.86E-06	7.8E-05	0	0	0	0	0	0	0	0	0	0		
28	aa	0.45E-05	0.000856	0.001732	0.000239	0.000286	0.001169	7.05E-05	0.000238	0.000464	0.003776	0.000538	0.002638	0.001984	0.000931	0.000358	0.000663	0.000518	0.001344	0.000848	0.000124	1.49E-05	0.001064	1.11E-05	0.000256	0.000633	5.57E-06	0	0	7.24E-05	1.86E-06	6.5E-05	
29	ab	0.000681	0.000529	0.001416	0.00036	0.000427	0.001797	7.61E-05	0.000455	0.001012	1.86E-06	0.002163	0.000173	0.000704	0.002278	0.000878	0.001043	0.000271	0.002181	0.000581	0.000232	4.27E-05	0.000121	0.000113	0.000494	0.000516	1.11E-05	0	0	0.000197	0.000626	0.000622	
30	ac	0	1.86E-06	7.43E-06	1.11E-05	5.57E-06	0	0	3.71E-06	0	1.3E-05	6.31E-05	0.0001	1.49E-05	5.75E-05	0	1.11E-05	5.94E-05	8.72E-05	0.002377	0	0	5.57E-06	1.86E-06	0	1.86E-06	0	0	0	1.86E-06	0	0	
31	ad	0.000145	0.00049	0.000343	0.000141	0.000572	7.98E-05	7.8E-05	0.000123	0.000533	0	0.000308	0.000128	0.000182	0.000386	0.000213	0.000525	0.000239	0.000496	0.000644	0.0001	2.41E-05	2.23E-05	6.87E-05	0.000234	0.000113	0.000518	0	0	4.08E-05	4.64E-05	5.2E-05	
32	ae	0.000329	0.000553	0.001669	0.000711	0.001403	0.000947	0.000718	0.000215	0.001043	2.23E-05	0.001194	0.00121	0.000724	0.002237	0.000867	0.000155	0.000308	0.000176	0.002231	0.000254	5.94E-05	0.000514	6.68E-05	0.000178	0							