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 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:

Завдання 1:

Для дослідження було обрано текст російською мовою «TEXT.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:

A	B	C
Параметр	**З пробілами**	**Без пробілів**
H1	4,383006841	4,468568723
H2 (step 1)	3,975230577	4,150663708
H2 (step 2)	3,974254627	4,149794329
R (H1)	0,123398632	0,098023485
R (H2 step 1)	0,204953885	0,162192322

Таблиці частот символів:

(З пробілом)

Символ	Частота
	0,162346
о	0,095222
а	0,070264
е	0,066722
и	0,055668
н	0,054587
т	0,047573
с	0,043704
л	0,042398
в	0,03856
р	0,038175
к	0,030044
д	0,025471
м	0,024757
у	0,024013
п	0,021519
я	0,019391
г	0,017368
ь	0,016753
ы	0,015897
з	0,014918
б	0,014467
ч	0,011416
й	0,009648
ж	0,008485
ш	0,00791
х	0,007145
ю	0,005433
ц	0,003388
э	0,002533
щ	0,002349
ф	0,001875

(Без пробілу)

Символ	Частота
о	0,113677
а	0,083882
е	0,079653
и	0,066457
н	0,065167
т	0,056793
с	0,052174
л	0,050615
в	0,046034
р	0,045574
к	0,035867
д	0,030408
м	0,029556
у	0,028667
п	0,025689
я	0,02315
г	0,020735
ь	0,02
ы	0,018979
з	0,017809
б	0,017271
ч	0,013629
й	0,011518
ж	0,01013
ш	0,009443
х	0,00853
ю	0,006486
ц	0,004045
э	0,003024
щ	0,002805
ф	0,002239

Частота біграм з перекриттям з пробілами:

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я
а	7,77E-06	0,000661	0,003407	0,000944	0,002102	0,000851	0,001186	0,00397	0,00012	0,000757	0,004262	0,008812	0,002665	0,005077	4,66E-06	0,000718	0,002668	0,003705	0,00462	0,000134	0,000504	0,000962	8,55E-05	0,000728	0,000969	0,000235	0	0	0	0,000773	0,002766	0,016571
б	0,001138	2,49E-05	3,73E-05	1,71E-05	1,4E-05	0,001965	4,66E-06	1,55E-06	0,000732	0	0,000246	0,000753	6,06E-05	0,000249	0,002155	0	0,001169	6,84E-05	7,77E-06	0,001141	0	4,66E-05	6,22E-06	2,49E-05	4,66E-06	0,000162	0,003541	0,000117	0	6,22E-06	0,000462	0,000313
в	0,005755	9,33E-06	2,18E-05	2,18E-05	0,000272	0,004159	0	0,000508	0,003258	0	0,000124	0,000927	0,000138	0,001339	0,006952	0,000305	0,000468	0,002959	0,000197	0,00079	0	3,89E-05	1,4E-05	2,95E-05	0,001236	6,22E-06	0,00264	0,000177	1,55E-06	0	0,000124	0,006088
г	0,001003	0	7,46E-05	0	0,00086	0,000555	0	0	0,000807	0	0,000104	0,001612	1,09E-05	0,000285	0,008917	0	0,001343	3,73E-05	3,11E-06	0,000771	0	0	0	2,18E-05	6,22E-06	0	0	0	0	0	0	0,000956
д	0,004005	3,42E-05	0,000911	9,33E-06	2,02E-05	0,004256	1,24E-05	4,66E-06	0,002709	0	0,000205	0,000608	8,71E-05	0,00163	0,003478	9,49E-05	0,002025	0,000264	0,000289	0,001595	1,24E-05	5,13E-05	0,000162	2,18E-05	0,000103	0	0,000653	0,000765	0	1,4E-05	0,000359	0,001092
е	9,33E-06	0,001039	0,001315	0,003323	0,002401	0,001437	0,000927	0,001078	0,000162	0,00228	0,000973	0,005913	0,003715	0,00666	0,000244	0,000933	0,006737	0,003416	0,003505	6,53E-05	1,4E-05	0,001028	0,000269	0,000886	0,000715	0,000642	0	0	0	0,000339	0,000367	0,01633
ж	0,001334	4,98E-05	0	2,33E-05	0,000622	0,00369	1,24E-05	0	0,00109	0	8,4E-05	1,55E-06	1,09E-05	0,001018	3,42E-05	0	0	2,02E-05	0	0,000188	0	0	0	4,66E-05	0	0	0	3,42E-05	0	2,02E-05	0	0,000205
з	0,005403	0,000143	0,000745	0,000429	0,000659	0,000207	0,000171	6,22E-06	0,000311	0	0,000162	0,000176	0,000255	0,001505	0,000684	0	0,000145	0,000146	1,09E-05	0,000257	0	0	3,11E-06	1,24E-05	1,24E-05	0	0,000529	0,001098	0	0,000112	0,000507	0,00123
и	7E-05	0,000482	0,00258	0,000474	0,001709	0,002325	0,000384	0,00164	0,000619	0,00141	0,001644	0,004876	0,002804	0,003163	0,000271	0,000215	0,000662	0,002771	0,003267	1,55E-06	1,55E-05	0,001662	0,001412	0,001162	0,000353	0,000104	0	0	0	0,000288	0,001564	0,017742
й	0	4,66E-06	0	0	0,000179	1,09E-05	0	0	0	0	3,89E-05	0,000224	4,35E-05	0,000336	1,24E-05	0	5,6E-05	0,000648	0,000267	0	0	0	3,58E-05	5,44E-05	8,09E-05	1,55E-06	0	0	0	0	0	0,007653
к	0,007478	0	0,000166	3,11E-06	1,87E-05	0,000473	1,09E-05	6,22E-06	0,002489	0	9,33E-06	0,000356	0	0,002256	0,008174	0	0,001707	6,69E-05	0,000375	0,001511	0	3,11E-06	1,09E-05	3,11E-06	6,22E-06	0	0	0	0	1,55E-05	0	0,004904
л	0,006375	2,33E-05	9,33E-06	0,000146	0,000327	0,003643	0,00033	6,22E-06	0,006296	0	0,000779	0,000114	4,66E-06	0,000249	0,006215	0,000103	0	0,001342	4,2E-05	0,001169	7,77E-06	0	0	0,000208	1,55E-06	1,55E-06	0,001208	0,00345	0	0,000818	0,001706	0,007845
м	0,002737	3,11E-06	0	5,91E-05	0	0,002685	0	3,11E-06	0,003427	0	6,69E-05	7,46E-05	1,55E-05	0,001106	0,003489	0,000222	4,2E-05	0,000131	1,55E-05	0,002164	1,55E-06	0	3,42E-05	2,64E-05	6,22E-06	1,55E-06	0,000614	2,33E-05	0	6,06E-05	0,000429	0,007319
н	0,010179	1,09E-05	9,33E-06	0,000165	0,001116	0,008616	2,33E-05	1,71E-05	0,00654	0	0,000218	3,11E-06	1,55E-06	0,002842	0,008294	9,33E-06	4,04E-05	0,000397	0,000589	0,002626	9,33E-06	0	0,000564	0,000188	3,58E-05	0,000143	0,003023	0,000992	0	0,000149	0,003323	0,004463
о	1,55E-06	0,003018	0,007949	0,004587	0,004413	0,00209	0,00158	0,000868	0,00077	0,003569	0,001656	0,006703	0,004802	0,006557	0,00016	0,001068	0,006116	0,00643	0,006389	4,82E-05	0,000435	0,000535	8,55E-05	0,001303	0,00128	0,000123	0	0	2,02E-05	0,000731	0,000614	0,021323
п	0,001151	0	0	1,4E-05	0	0,002415	0	0	0,000743	0	6,06E-05	0,000704	0	5,13E-05	0,008476	7,93E-05	0,005901	3,11E-06	4,82E-05	0,000678	1,55E-06	0	1,55E-06	7,77E-06	1,24E-05	0	0,000194	0,000583	0	0	0,000353	4,04E-05
р	0,008285	0,000121	0,000255	0,000274	0,000397	0,00591	0,000364	3,27E-05	0,005169	0	0,000252	6,06E-05	0,00023	0,000638	0,006801	4,2E-05	4,2E-05	0,000235	0,000636	0,003037	2,02E-05	0,000131	9,49E-05	7,93E-05	0,000232	6,38E-05	0,001421	0,000796	1,55E-06	0,000204	0,001051	0,001301
с	0,001396	4,2E-05	0,001698	7,77E-06	0,000303	0,00332	2,18E-05	4,98E-05	0,001642	0	0,004631	0,002287	0,000967	0,000701	0,002437	0,001672	0,000257	0,000743	0,010358	0,000715	1,71E-05	0,000182	2,02E-05	0,000317	4,98E-05	0	0,000395	0,003292	0	0,000128	0,00302	0,003038
т	0,005344	2,8E-05	0,002094	9,33E-06	0,000129	0,004531	1,55E-06	6,22E-06	0,003332	0	0,000342	0,000269	4,66E-05	0,000869	0,013195	5,91E-05	0,00279	0,000802	0,000104	0,001742	1,4E-05	2,18E-05	0,000118	0,00016	0	1,87E-05	0,001532	0,004904	4,66E-06	5,91E-05	0,000487	0,004559
у	1,87E-05	0,000502	0,00083	0,001222	0,001905	0,000146	0,001249	0,000855	1,71E-05	9,02E-05	0,000972	0,001482	0,000976	0,000295	1,24E-05	0,000543	0,000566	0,001382	0,001385	1,55E-06	6,22E-06	0,000355	1,55E-06	0,000687	0,000751	0,000191	0	0	2,18E-05	0,000992	6,69E-05	0,006489
ф	0,000177	0	0	0	0	0,000137	0	0	0,000627	0	1,09E-05	9,17E-05	4,66E-06	0	5,75E-05	0	0,000476	6,22E-06	6,22E-06	7E-05	4,66E-06	0	0	0	0	0	1,87E-05	1,55E-06	0	0	0	0,000187
х	0,000964	0	0,000128	0	0	3,89E-05	0	0	0,000152	0	0	6,06E-05	5,44E-05	0,000148	0,002236	1,55E-06	7,31E-05	3,73E-05	2,02E-05	0,00012	0	0	0	0	0	0	0	3,11E-06	0	0	0	0,003108
ц	0,000561	0	1,87E-05	2,64E-05	0	0,00107	0	0	0,000222	0	0,000171	4,66E-06	0	2,49E-05	0,000409	0	0	0	0	0,000485	0	0	0	0	0	0	0,000117	0	0	0	0	0,000278
ч	0,002108	0	0	0	0	0,002934	0	0	0,00116	0	0,000222	3,58E-05	0	0,000456	9,33E-05	0	2,02E-05	0	0,003393	0,000569	0	0	0	0	7E-05	0	0	0,000193	1,55E-06	0	0	0,00016
ш	0,001362	0	1,55E-05	0	0	0,002152	0	0	0,001906	0	0,000407	0,000452	1,87E-05	0,000328	0,000327	5,44E-05	3,11E-06	1,55E-06	0,000159	0,000285	0	0	0	0	0	0	0	0,00032	0	0	0	0,000118
щ	0,000404	0	0	0	0	0,001176	0	0	0,000644	0	0	0	0	2,8E-05	0	0	1,55E-06	0	0	6,53E-05	0	0	0	0	0	0	0	1,87E-05	0	0	0	1,24E-05
ы	0	0,000494	0,00106	0,00011	7,93E-05	0,000888	2,8E-05	4,66E-05	1,24E-05	0,001487	0,00018	0,002171	0,001445	0,000344	1,55E-06	0,000123	0,000345	0,00077	0,000529	4,66E-06	0	0,000849	1,55E-06	0,000129	0,000515	1,55E-06	0	0	0	0	0	0,004284
ь	0	0,000104	4,66E-06	4,04E-05	3,42E-05	0,001137	0	0,000129	7,31E-05	0	0,001085	0	0,000337	0,001001	4,35E-05	0	0	0,000871	0,000101	0	6,22E-06	0	8,55E-05	4,04E-05	0,000403	7,77E-06	0	0	0	0,000522	0,000428	0,010298
э	0	0	0	7,77E-06	1,55E-06	0	0	1,55E-06	0	7,77E-06	5,13E-05	8,24E-05	7,77E-06	4,66E-05	0	9,33E-06	4,51E-05	7E-05	0,002152	0	0	4,66E-06	1,55E-06	0	1,55E-06	0	0	0	0	0	0	4,2E-05
ю	0	0,000302	0	1,55E-06	0,000344	0	6,22E-06	2,64E-05	0	0	3,11E-06	3,89E-05	2,49E-05	8,09E-05	0	0	6,53E-05	0,000141	0,000412	0	0	6,22E-06	4,51E-05	5,75E-05	7,15E-05	0,000429	0	0	0	3,58E-05	0	0,003342
я	0	3,27E-05	0,000263	0,000295	0,000707	9,49E-05	0,000507	0,001521	1,09E-05	1,87E-05	7,46E-05	0,000872	0,000323	0,000599	0	1,4E-05	3,89E-05	0,00086	0,001379	0	1,55E-06	0,000138	3,27E-05	0,000118	3,11E-06	0,000145	0	0	0	8,55E-05	0,000199	0,011057
	0,002996	0,007338	0,014968	0,005158	0,006859	0,003811	0,001667	0,004141	0,010628	2,95E-05	0,011009	0,002632	0,005708	0,014706	0,012048	0,015254	0,004372	0,01538	0,007314	0,003778	0,000804	0,00113	0,000303	0,005103	0,000992	7,31E-05	1,24E-05	4,66E-06	0,00248	8,09E-05	0,001566	0

Без пробілів:

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
а	0,000388	0,001583	0,005821	0,001719	0,003269	0,001533	0,001533	0,005192	0,001572	0,000906	0,006516	0,010869	0,004093	0,007893	0,001472	0,002712	0,003733	0,00633	0,006404	0,000555	0,000713	0,001277	0,000158	0,001298	0,001275	0,00028	0	0	0,000347	0,000928	0,003512
б	0,001368	3,53E-05	5,94E-05	2,23E-05	2,04E-05	0,002359	9,28E-06	3,71E-06	0,000906	0	0,000303	0,000898	8,17E-05	0,000316	0,002703	5,57E-06	0,001426	9,65E-05	1,11E-05	0,001383	0	5,75E-05	7,43E-06	3,34E-05	7,43E-06	0,000193	0,004227	0,000139	3,34E-05	7,43E-06	0,000559
в	0,006987	0,000249	0,000464	0,000412	0,000707	0,005125	2,97E-05	0,000759	0,004199	0	0,000733	0,001296	0,000349	0,002116	0,008764	0,001062	0,000802	0,004247	0,000796	0,001088	1,67E-05	7,98E-05	7,24E-05	0,000239	0,001548	9,28E-06	0,003154	0,000212	0,00034	5,57E-06	0,000173
г	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
д	0,004798	9,47E-05	0,001175	9,28E-05	5,57E-05	0,00511	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	0,00078	0,000913	2,04E-05	1,86E-05	0,00044
е	0,000195	0,002196	0,003445	0,004544	0,003724	0,002047	0,001337	0,001964	0,001448	0,002721	0,002189	0,007455	0,005534	0,009441	0,001611	0,003139	0,008539	0,006048	0,005053	0,000583	6,87E-05	0,001427	0,000334	0,001561	0,000963	0,000772	0	0	0,00029	0,000414	0,000611
ж	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
з	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
и	0,000388	0,001426	0,00525	0,001285	0,002979	0,003271	0,000613	0,002545	0,001992	0,001689	0,003195	0,006128	0,003937	0,005701	0,002072	0,002562	0,001437	0,005352	0,004713	0,000555	0,000145	0,002138	0,00173	0,00197	0,000587	0,000137	0	3,71E-06	0,000256	0,000349	0,002055
й	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	0,00046	0,000967	0,000626	0,000848	0,000416	0,001736	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
к	0,009023	0,000644	0,000622	0,00021	0,000299	0,000648	0,000167	9,47E-05	0,003297	1,86E-06	0,000449	0,000497	0,00023	0,003219	0,010191	0,000394	0,002135	0,000624	0,000726	0,001914	2,6E-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
л	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	0,002374	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	0,002073
м	0,003351	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
н	0,012224	0,000343	0,000553	0,000343	0,001533	0,010356	5,94E-05	0,000212	0,008182	0	0,00052	3,9E-05	0,000124	0,003969	0,010286	0,000681	0,001169	0,000999	0,000923	0,003326	2,41E-05	5,57E-05	0,000679	0,000317	6,5E-05	0,000173	0,003609	0,001184	3,53E-05	0,00018	0,004002
о	0,00023	0,005042	0,012032	0,006155	0,00628	0,003113	0,002304	0,001658	0,002294	0,00426	0,003223	0,008372	0,006562	0,01043	0,002344	0,003542	0,00788	0,009907	0,00923	0,005598	0,000692	0,000794	0,000149	0,002599	0,00165	0,000158	0	1,86E-06	0,000408	0,000882	0,001088
п	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
р	0,009937	0,000208	0,000436	0,000368	0,00052	0,007076	0,000438	6,68E-05	0,006284	1,86E-06	0,000408	8,91E-05	0,000306	0,000898	0,008233	0,000243	9,28E-05	0,000477	0,000806	0,003688	4,83E-05	0,000174	0,000115	0,000115	0,000282	7,61E-05	0,001698	0,00095	1,86E-05	0,000247	0,001272
с	0,001736	0,000213	0,002274	0,000134	0,000512	0,004034	0,000106	0,000124	0,002064	1,86E-06	0,005777	0,002809	0,001247	0,001227	0,003087	0,002307	0,000446	0,001273	0,01269	0,00103	5,2E-05	0,00023	2,78E-05	0,000429	9,28E-05	1,86E-06	0,000471	0,00393	6,87E-05	0,000152	0,003627
т	0,006491	0,000321	0,002992	0,00016	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	0,005855	0,000145	7,24E-05	0,000655
у	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
ф	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	1,86E-06	0	5,57E-06
х	0,001194	0,000165	0,000484	0,000173	0,000152	0,000106	4,27E-05	7,98E-05	0,000483	1,86E-06	0,000251	0,000217	0,000208	0,000408	0,002931	0,000392	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
ц	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	3,71E-06	0	0	0,000139	0	3,71E-06	0	3,71E-06
ч	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
ш	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	8,17E-05	5,57E-06	1,3E-05	0,000191	0,00034	1,86E-06	0	1,86E-06	5,57E-06	5,57E-06	0	0	0,000382	0	0	0
щ	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	2,23E-05	1,86E-06	0	0
ы	7,43E-05	0,000856	0,001732	0,000239	0,000286	0,001169	7,05E-05	0,000238	0,000464	0,001776	0,000538	0,002638	0,001984	0,000932	0,000358	0,000663	0,000518	0,001344	0,000848	0,000124	1,49E-05	0,001066	1,11E-05	0,000256	0,000633	5,57E-06	0	0	7,24E-05	1,86E-06	6,5E-05
ь	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
э	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	0,002577	0	0	5,57E-06	1,86E-06	0	1,86E-06	0	0	0	1,86E-06	0	0
ю	0,000145	0,00049	0,000343	0,000141	0,000572	7,98E-05	7,8E-05	0,000123	0,000353	0	0,000308	0,000128	0,000182	0,000386	0,00023	0,000325	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	4,46E-05	4,64E-05	5,2E-05
я	0,000329	0,000553	0,001669	0,000711	0,001403	0,00047	0,000718	0,00215	0,001043	2,23E-05	0,001194	0,00121	0,000724	0,002257	0,000867	0,001155	0,000308	0,002176	0,002231	0,000254	5,2E-05	0,000234	5,94E-05	0,000514	6,68E-05	0,000178	0	0	0,000158	0,000111	0,000332

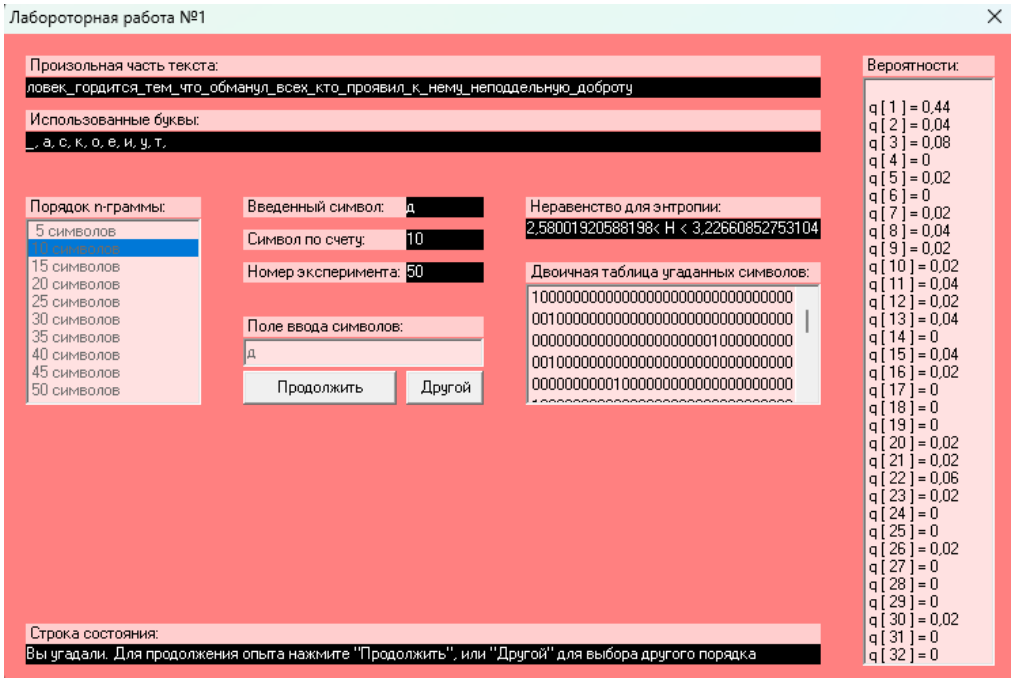
Завдання 2:

Для проведення експерименту використовувалася програма CoolPinkProgram та очищений російський текст.

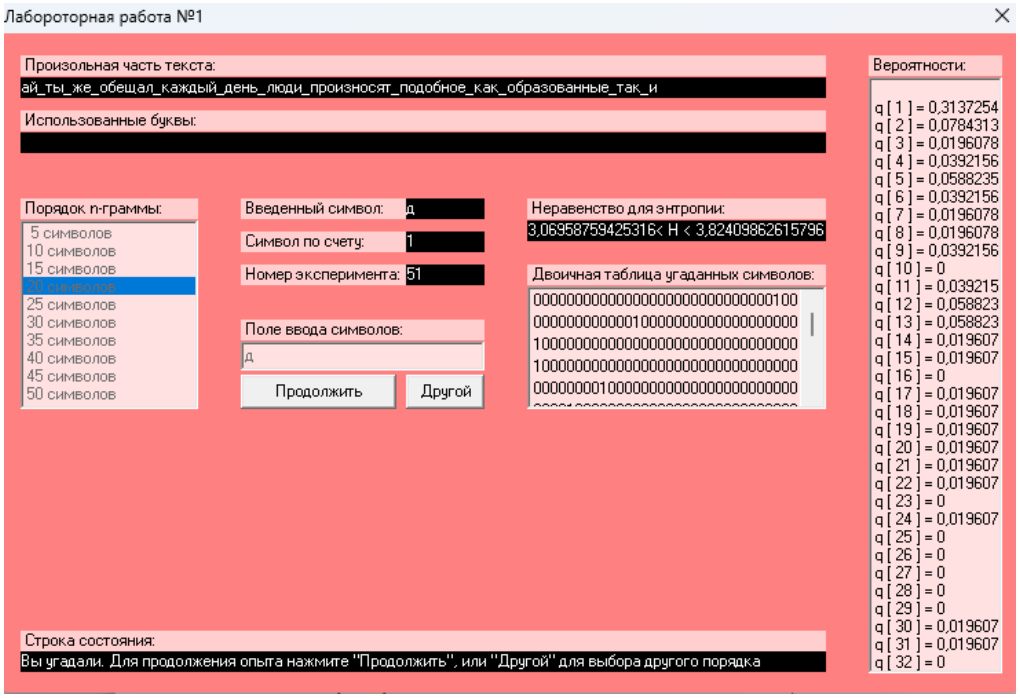
Було проведено по 50 експериментів для кожного порядку n-грами: n=10, 20, 30.

Експерименти із CoolPinkProgram:

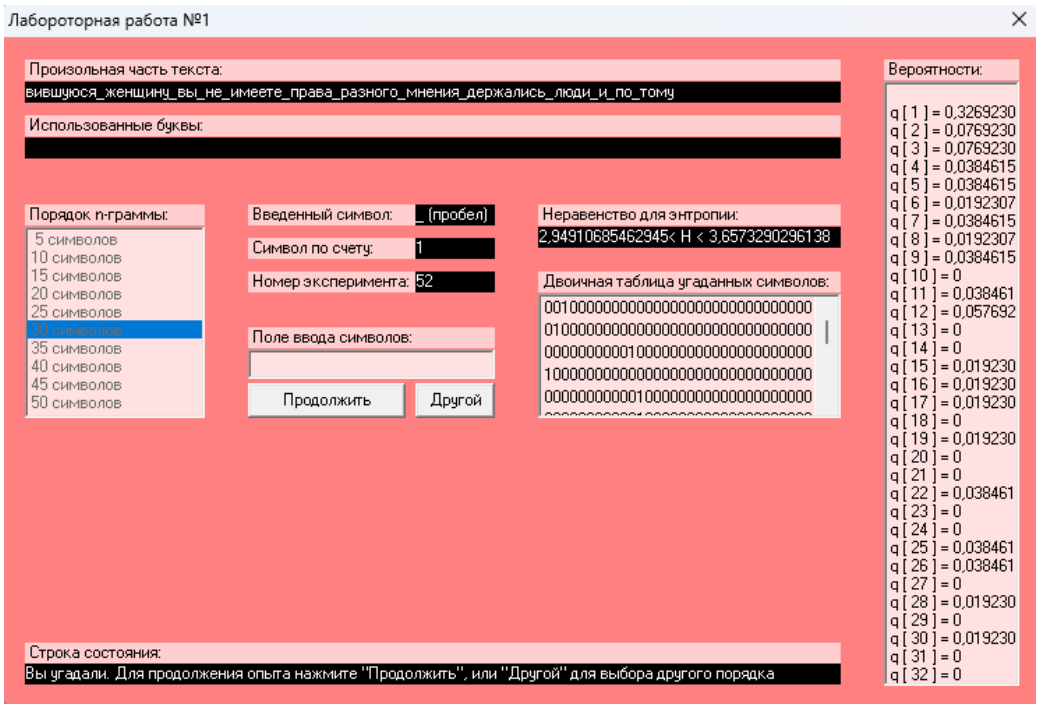
H10



H20



H30



Завдання 3:

$H_0 = \log_2 32 = 5$

$R = 1 - H/H_0$

$0.3547 < R^{(10)} < 0.4840$

$0.2352 < R^{(20)} < 0.3861$

$0.2685 < R^{(30)} < 0.4102$

Висновок:

Під час виконання комп’ютерного практикуму були опановані методи оцінки ентропії та надлишковості мови. За допомогою програми, реалізованої мовою Python, було проаналізовано великий масив російськомовного тексту. Отримані результати показали, що питома ентропія на символ для моделі біграм H_2 (3.9752) є меншою за ентропію H_2 (4.3830), оскільки біграмна модель враховує статистичні зв’язки між сусідніми літерами. Порівняння результатів для тексту з пробілами та без них показало, що наявність пробілу, як найбільш частотного символу, вносить додаткову впорядкованість і знижує невизначеність. Другий етап роботи з програмою CoolPinkProgram дозволив оцінити умовну ентропію на основі суб'єктивного вгадування символів. Експерименти для $H^{(10)}$, $H^{(20)}$ та $H^{(30)}$ підтвердили, що чим довший контекст ми знаємо, тим легше передбачити наступну літеру, хоча через людський фактор оцінки виявилися дещо вищими за теоретично можливі.