



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

Виконали:
студенти групи ФБ-32
Кошеленко Н.Е.
Кухарук І.А.

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 $H^{(10)}$, $H^{(20)}$, $H^{(30)}$.

3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Для виконання даної лабораторної роботи було взято текст художнього твору «Приключення Оливера Твіста» (автор Чарльз Дікенс).

На першому етапі виконано попередню фільтрацію тексту за допомогою програмного коду, згідно з методичними вказівками. У результаті отримано очищений текст, який відповідає необхідним умовам:

- усі символи, окрім текстових (літер та пробілів), вилучено або замінено на пробіли;
- усі прописні літери замінено на відповідні рядкові;
- послідовності пробілів та розділових знаків (зокрема символів кінця рядка) замінено одним пробілом, або вилучено, якщо пробіл не входить до алфавіту;
- буква «ё» замінена на «е», а «ъ» - на «ь» для уніфікації алфавіту.

Отриманий текст збережено у двох варіантах:

1. із пробілами, де пробіл входить до алфавіту ($m = 32$);
2. без пробілів, де пробіл повністю вилучено ($m = 31$).

Частоти монограм з пробілом та без пробілів.

Літера	Кількість	Частота	Літера	Кількість	Частота
	130039	0.161705	о	74523	0.110546
о	74523	0.09267	е	56902	0.084407
е	56902	0.070758	а	51583	0.076517
а	51583	0.064144	и	47076	0.069832
и	47076	0.058539	н	41486	0.061539
н	41486	0.051588	т	41031	0.060864
т	41031	0.051022	с	37542	0.055689
с	37542	0.046684	л	36180	0.053669
л	36180	0.04499	в	29716	0.04408
в	29716	0.036952	р	29196	0.043309
р	29196	0.036305	м	23268	0.034515
м	23268	0.028934	к	23181	0.034386
к	23181	0.028826	д	21188	0.03143
д	21188	0.026347	у	19815	0.029393
у	19815	0.02464	п	17450	0.025885
п	17450	0.021699	ь	13436	0.019931
ь	13436	0.016708	я	13260	0.01967
я	13260	0.016489	б	12826	0.019026
б	12826	0.015949	ы	12707	0.018849
ы	12707	0.015801	г	11939	0.01771
г	11939	0.014846	з	11507	0.017069
з	11507	0.014309	ч	9907	0.014696
ч	9907	0.012319	й	8578	0.012724
й	8578	0.010667	ж	7041	0.010444
ж	7041	0.008756	ш	5862	0.008696
ш	5862	0.007289	х	5353	0.007941
х	5353	0.006657	ю	3659	0.005428
ю	3659	0.00455	э	2950	0.004376
э	2950	0.003668	щ	2259	0.003351
щ	2259	0.002809	ц	1842	0.002732
ц	1842	0.002291	ф	874	0.001296
ф	874	0.001087			

Біграми з пробілами та перекриттям (with_sp_bi_overlap).

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф
	0.001293	0.007684	0.013823	0.003228	0.007567	0.004538	0.001592	0.004503	0.009273	1.24e-06	0.008795	0.002676	0.00838	0.014732	0.012706	0.015484	0.003678	0.016877	0.00718	0.004533	0.008007
а	0.016052	0	0.000696	0.002858	0.000777	0.001762	0.000959	0.000885	0.003798	0.000132	0.001358	0.004554	0.000783	0.003111	0.002674	6.22e-06	0.000586	0.002257	0.003471	0.003974	0.000399
б	0.000251	0.00122	6.22e-06	6.84e-05	1.24e-06	2.36e-05	0.002155	6.22e-06	1.24e-06	0.000939	0	9.33e-05	0.001245	6.71e-05	0.000229	0.002232	0	0.001409	0.000148	3.73e-06	0.001159
в	0.005388	0.005328	6.22e-06	3.85e-05	1.24e-06	0.000279	0.006256	0	0.000607	0.002716	0	0.000213	0.000749	0.000928	0.005988	0.000284	0.000939	0.001881	0.000238	0.000817	
г	0.000722	0.000939	0	7.46e-06	0	0.001395	0.000149	0	0.000608	0	9.82e-05	0.001368	9.95e-06	0.000313	0.000745	0	0.000981	2.61e-05	1.49e-05	0.000502	
д	0.001176	0.004233	3.98e-05	0.000994	1.99e-05	3.11e-05	0.004451	0.001109	0.000102	0.002179	0	0.000195	0.000813	8.08e-05	0.001806	0.004152	6.09e-05	0.000972	0.000338	0.000256	
е	0.016511	1.24e-05	0.001128	0.001673	0.0033	0.003433	0.001384	0.000842	0.001084	0.00106	0.002136	0.001015	0.005204	0.003712	0.007485	0.000147	0.000791	0.007955	0.004306	0.004794	
ж	0.000129	0.001398	1.62e-05	0	6.22e-06	0.000802	0.003397	3.73e-06	0	0.001712	0	0.000102	8.70e-06	1.24e-06	0.000868	3.11e-05	0	4.97e-06	1.12e-05	1.24e-06	
з	0.001476	0.005751	0.000153	0.000812	0.000388	0.000807	0.000213	4.97e-05	2.61e-05	0.000499	0	0.000157	0.000251	0.000282	0.00162	0.000399	0	0.000372	1.24e-06	1.49e-05	
и	0.018057	4.23e-05	0.00049	0.00386	0.00046	0.001567	0.00217	0.000235	0.001798	0.00045	0.000904	0.002167	0.000578	0.002728	0.003052	0.000109	0.000165	0.000629	0.00464	0.003906	
й	0.008079	0	0	6.22e-06	4.97e-06	0.000189	1.62e-05	0	3.73e-06	0	0	0.000505	0.00028	4.10e-05	0.000333	4.97e-06	6.22e-05	1.24e-06	0.0003	0.000561	
к	0.004409	0.007273	0	0.000101	0	0	0.000395	4.73e-05	1.24e-06	0.002393	0	7.46e-06	0.000929	0	0.000653	0.00058	0	0.001761	0.000662	0.000561	
л	0.008115	0.005746	2.86e-05	0.000148	6.34e-05	0.004106	0.000357	1.12e-05	0.000201	0	0.000398	0.000178	1.62e-05	0.000306	0.005981	9.33e-05	0	0.001728	0.000114	0.00166	
м	0.008021	0.002519	0.00051	0	3.11e-05	1.24e-06	0.003786	0	3.73e-06	0.004509	0	0.000116	0.000159	1.99e-05	0.001482	0.003714	7.34e-05	9.70e-05	8.83e-05	1.12e-05	
н	0.003937	0.008632	6.22e-06	9.95e-06	0.000159	0.000165	0.008537	2.49e-06	1.12e-05	0.007452	0	0.000407	0.000223	0	0.002262	0.008677	0	3.98e-05	0.000453	0.000814	
о	0.019635	0	0.003918	0.005882	0.004718	0.004469	0.001895	0.001854	0.001488	0.00075	0.003987	0.002235	0.00677	0.005639	0.005585	0.000154	0.001159	0.004904	0.000601	0.000758	
п	3.23e-05	0.001023	0	0	0	0	0.001825	0	0	0.000772	0	0.00109	0.000333	0.00077	0.000144	0.000807	1.87e-05	0.00647	1.24e-06	9.45e-05	
р	0.002625	0.006949	0.000157	0.000276	0.000144	0.000368	0.005616	0.000384	7.34e-05	0.00459	0	0.000218	0.000286	0.000192	0.000962	0.006638	0.000126	7.46e-05	0.000155	0.00051	
с	0.004039	0.001813	0.000112	0.001566	1.62e-05	0.000271	0.002742	4.60e-05	7.46e-06	0.001911	0	0.003586	0.003489	0.000802	0.001027	0.002335	0.001728	0.000141	0.000814	0.011088	
т	0.005075	0.005292	3.11e-05	0.002634	3.73e-06	0.000107	0.006693	0	3.73e-06	0.003867	0	0.000659	0.000856	2.36e-05	0.001101	0.012296	0.000153	0.002297	0.001114	0.001212	
у	0.006722	3.11e-05	0.000485	0.000749	0.001161	0.002043	0.00029	0.000125	0.000255	0.000162	0.00118	0.000914	0.000822	0.001061	0.000532	1.12e-05	0.00064	0.000279	0.001195	0.001568	
ф	4.23e-05	9.33e-05	0	0	0	0.000466	0	0	0.000163	0	0	9.95e-06	1.24e-06	0	9.70e-05	0	3.23e-05	6.22e-06	1.24e-06	9.20e-05	
х	0.003027	0.000471	0	0.000219	0	0	0	0.000216	0	0	0.000216	0	0.00015	5.97e-05	0.000175	0.000198	1.24e-06	0.000128	3.73e-05	2.49e-06	
ц	0.000311	0.000567	0	4.60e-05	0	0.000741	0	9.58e-05	0	1.24e-05	1.24e-06	1.24e-06	0	0.000224	0	0	0	0	0	0.000123	
ч	6.59e-05	0.002121	0	6.22e-06	0	0.003127	0	0	0.001995	0	0.000234	3.48e-05	2.49e-06	0.000612	7.34e-05	0	3.36e-05	0	0.003909	0.000527	
ш	5.60e-05	0.000966	0	2.49e-05	0	0.002151	0	0.001811	0	0.000669	0.00055	3.36e-05	0.000199	0.000164	1.49e-05	2.49e-06	0	5.84e-05	0.000264	0	
щ	3.66e-05	0.000498	0	0	0	0.000419	0	0.000285	0	0.000285	0	0	0	0	0	2.49e-06	0	2.49e-06	0	0	

	а	б	в	г	д	е	ж	з	и	й	к	л
	0.001293	0.007684	0.013823	0.003228	0.007567	0.004538	0.001592	0.004503	0.009273	1.24e-06	0.008795	0.002676
а	0.016052	0	0.000696	0.002858	0.000777	0.001762	0.000959	0.000885	0.003798	0.000132	0.001358	0.004554
б	0.000251	0.00122	6.22e-06	6.84e-05	1.24e-06	2.36e-05	0.002155	6.22e-06	1.24e-06	0.000939	0	9.33e-05
в	0.005388	0.005328	6.22e-06	3.85e-05	1.24e-06	0.000279	0.006256	0	0.000607	0.002716	0	0.000607
г	0.000722	0.000939	0	7.46e-06	0	0.001395	0.000149	0	0	0.000608	0	9.82e-05
д	0.001176	0.004233	3.98e-05	0.000994	1.99e-05	0.004451	0.001109	0.000102	0.002179	0	0.000195	0.000813
е	0.016511	1.24e-05	0.001128	0.001673	0.0033	0.003433	0.001384	0.000842	0.001084	0.000106	0.002136	0.001015
ж	0.000129	0.001398	1.62e-05	0	6.22e-06	0.000802	0.003397	3.73e-06	0	0.001712	0	0.000102
з	0.001476	0.005751	0.000153	0.000812	0.000388	0.000807	0.000213	4.97e-05	2.61e-05	0.000499	0	0.000157
и	0.018057	4.23e-05	0.00049	0.00386	0.00046	0.001567	0.00217	0.000235	0.001798	0.00045	0.000904	0.002167
й	0.008079	0	0	6.22e-06	4.97e-06	0.000189	1.62e-05	0	3.73e-06	0	0	0.000505
к	0.004409	0.007273	0	0.000101	0	0	0.000395	4.73e-05	1.24e-06	0.002393	0	7.46e-06
л	0.008115	0.005746	2.86e-05	0.000148	6.34e-05	0.004106	0.000357	1.12e-05	0.000201	0	0.000398	0.000178
м	0.008021	0.002519	0.00051	0	3.11e-05	1.24e-06	0.003786	0	3.73e-06	0.004509	0	0.000116
н	0.003937	0.008632	6.22e-06	9.95e-06	0.000159	0.000165	0.008537	2.49e-06	1.12e-05	0.007452	0	0.000407
о	0.019635	0	0.003918	0.005882	0.004718	0.004469	0.001895	0.001854	0.001488	0.00075	0.003987	0.002235
п	3.23e-05	0.001023	0	0	0	0	0.001825	0	0	0.000772	0	0.000109
р	0.002625	0.006949	0.000157	0.000276	0.000144	0.000368	0.005616	0.000384	7.34e-05	0.00459	0	0.000218
с	0.004039	0.001813	0.000112	0.001566	1.62e-05	0.000271	0.002742	4.60e-05	7.46e-06	0.001911	0	0.003586
т	0.005075	0.005292	3.11e-05	0.002634	3.73e-06	0.000107	0.006693	0	3.73e-06	0.003867	0	0.000659
у	0.006722	3.11e-05	0.000485	0.000749	0.001161	0.002043	0.00029	0.001125	0.000255	0.000162	0.000118	0.000914
ф	4.23e-05	9.33e-05	0	0	0	0.000466	0	0	0.000163	0	0	9.95e-06
х	0.003027	0.000471	0	0.000219	0	0	0.000216	0	0	0.000216	0	0
ц	0.000311	0.000567	0	4.60e-05	0	0	0.000741	0	0	9.58e-05	0	1.24e-05
ч	6.59e-05	0.002121	0	6.22e-06	0	0	0.003127	0	0	0.001995	0	0.000234
ш	5.60e-05	0.000966	0	2.49e-05	0	0	0.002151	0	0	0.001811	0	0.000669

Біграми без пробілів та з перекриттям (no_sp_bi_overlap).

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х
а	0.000177	0.001688	0.004907	0.001307	0.003008	0.00174	0.001218	0.005069	0.00124	0.00162	0.006485	0.009663	0.00475	0.005202	0.001558	0.002421	0.003179	0.006058	0.005558	0.001004	0.000178	0.001104
б	0.001457	8.90e-06	9.35e-05	1.48e-06	3.41e-05	0.002575	7.42e-06	5.93e-06	0.001147	0	0.000117	0.001488	8.90e-05	0.000285	0.002701	1.04e-05	0.001682	0.00019	7.42e-06	0.0014	0	6.97e-05
в	0.006383	0.000223	0.000359	0.000231	0.000708	0.007617	5.93e-05	0.000847	0.003496	0	0.000899	0.001052	0.000292	0.001535	0.007555	0.000942	0.001329	0.002965	0.000869	0.001113	9.90e-06	6.53e-05
г	0.00113	6.82e-05	7.27e-05	1.19e-05	0.001706	0.000208	7.42e-06	2.67e-05	0.000785	0	0.000183	0.001635	4.75e-05	0.00046	0.009178	9.35e-05	0.	0.00138	4.15e-05	0.000629	0.	4.45e-06
д	0.005073	8.31e-05	0.001268	6.53e-05	8.01e-05	0.005346	0.001335	0.003187	0.00271	0	0.000313	0.001026	0.00016	0.002347	0.005038	0.000181	0.001196	0.000953	0.000347	0.001715	7.42e-06	9.20e-05
е	0.000138	0.002366	0.003658	0.00432	0.004919	0.00111	0.001188	0.002	0.00129	0.002548	0.002124	0.000564	0.005677	0.010452	0.001685	0.002968	0.000992	0.002702	0.000645	0.000865	9.97e-05	0.000712
ж	0.001669	2.52e-05	8.90e-06	1.19e-05	0.000901	0.00046	5.93e-06	1.48e-06	0.002047	0	0.000132	1.34e-05	5.93e-06	0.001062	4.90e-05	1.48e-05	5.93e-06	2.37e-05	1.34e-05	0.000181	0.	2.97e-06
з	0.006875	0.000372	0.00113	0.000516	0.001096	0.002089	6.97e-05	0.00101	0.000654	0	0.000362	0.000322	0.000411	0.002078	0.000584	0.000177	0.000497	0.000474	6.94e-05	0.00053	7.42e-06	1.48e-06
и	0.000221	0.001698	0.006659	0.000972	0.002796	0.003203	0.000452	0.002836	0.001807	0.001078	0.003708	0.006935	0.004198	0.005402	0.00198	0.002554	0.001222	0.000828	0.005563	0.000675	6.53e-05	0.000213
й	9.05e-05	0.001045	0.000745	0.000208	0.000909	0.000169	0.001101	0.000617	0.	0.	0.001167	0.000546	0.000749	0.001083	0.000366	0.000978	0.000323	0.001423	0.000992	0.000223	6.58e-05	6.09e-05
к	0.008718	0.000331	0.000544	8.75e-05	0.0003	0.00057	0.000159	9.20e-05	0.003139	0.	0.000399	0.001193	0.000233	0.001193	0.010011	0.000424	0.002222	0.001307	0.001019	0.001967	1.78e-05	4.30e-05
л	0.006923	0.00046	0.000762	0.000478	0.00055	0.000534	0.000458	0.000223	0.00246	0.	0.000863	0.000279	0.000923	0.001154	0.000794	0.000801	0.000181	0.001111	0.000491	0.002198	0.000199	4.30e-05
м	0.003103	0.000918	0.000682	0.00026	0.000501	0.004687	0.000141	0.000203	0.000609	0.	0.000666	0.000323	0.000405	0.00256	0.005193	0.001083	0.00034	0.001206	0.000362	0.002817	7.86e-05	8.01e-05
н	0.003539	0.000255	0.000494	0.000263	0.000335	0.002671	2.82e-05	0.000196	0.009204	0.	0.000699	0.000347	0.000157	0.001318	0.001074	0.000626	0.00016	0.001104	0.001124	0.003817	1.44e-05	6.38e-05
о	0.000117	0.005836	0.009133	0.00596	0.006396	0.000126	0.002502	0.002534	0.001934	0.004757	0.003575	0.000862	0.007902	0.008832	0.002271	0.003821	0.000395	0.000195	0.009593	0.001038	0.000148	0.000761
п	0.001221	1.48e-06	1.48e-06	1.48e-06	1.48e-06	0.001738	1.48e-06	1.48e-06	0.000927	0.	0.000134	0.000994	0.	0.000179	0.01051	2.34e-05	0.00793	4.55e-06	0.000114	0.000935	0.	0.5
р	0.008326	0.000957	0.00049	0.000313	0.000616	0.002672	0.000466	0.000132	0.005576	0.	0.000398	0.000148	0.000289	0.001323	0.00812	0.000386	0.00012	0.000252	0.000702	0.002866	8.01e-05	0.000141
с	0.002185	0.000383	0.002253	0.000129	0.00049	0.00361	0.000141	0.000111	0.002428	0.	0.004618	0.000231	0.0014	0.00181	0.003563	0.002502	0.000263	0.000368	0.01356	0.001078	3.56e-05	0.000322
т	0.006838	0.0003	0.000371	8.90e-05	0.000383	0.002803	5.34e-05	0.000157	0.004965	0.00111	0.00111	0.002432	0.001924	0.001521	0.000681	0.002866	0.000194	0.000395	0.001747	0.00145	9.20e-06	4.4
у	0.000157	0.000813	0.001556	0.0001541	0.002815	0.000524	0.001439	0.000473	0.000848	0.000141	0.001604	0.002255	0.001636	0.001644	0.000561	0.001472	0.000408	0.002215	0.002185	0.000163	5.79e-05	0.000576
ф	0.000113	0.1	0.34e-05	0.	4.54e-06	0.000568	0.	0.	0.000199	0.	0.1	1.19e-05	2.97e-06	4.45e-06	0.000119	2.97e-06	4.01e-05	1.78e-05	1.48e-06	0.0001	3.26e-05	0
х	0.000586	0.000179	0.000521	6.23e-05	0.000249	9.64e-05	3.71e-05	8.60e-05	0.0005	0.	0.000203	0.0003	0.000251	0.00046	0.002593	0.000383	0.000243	0.000448	0.000139	0.000283	3.71e-05	3.26e-05
ц	0.000681	8.90e-06	8.15e-05	2.97e-06	1.48e-05	0.000897	1.48e-06	2.97e-06	0.000138	0.	5.34e-05	2.97e-06	1.48e-05	3.26e-05	0.002322	0.01e-05	4.45e-06	2.37e-05	1.63e-05	0.000159	1.48e-06	4.45e-06
ч	0.002531	1.48e-06	1.19e-05	1.48e-06	2.97e-06	0.000732	0.2	97e-06	0.002394	0.	0.000282	4.15e-05	5.93e-06	0.000737	9.64e-05	4.45e-06	4.01e-05	1.19e-05	0.003664	0.00083	1.48e-06	0
ш	0.000154	5.93e-06	3.26e-05	1.48e-06	0.	0.002568	0.	0.	0.002169	0.	0.000798	0.000656	4.75e-05	0.000243	0.002023	2.52e-05	7.42e-06	5.93e-06	6.97e-05	0.000316	0.	2.97e-06
щ	0.000512	1.48e-06	1.48e-06	0.	0.	0.001413	0.	0.	0.001111	0.	1.48e-06	0.	0.	0.8	6.06e-05	8.90e-06	2.97e-06	2.97e-06	1.48e-06	9.49e-05	0.	1.48e-06

	а	б	в	г	д	е	ж	з	и	й	к	л
а	0.000177	0.001688	0.004907	0.001307	0.003008	0.00174	0.001218	0.005069	0.00124	0.00162	0.006485	0.009663
б	0.001457	8.90e-06	9.35e-05	1.48e-06	3.41e-05	0.002575	7.42e-06	5.93e-06	0.001147	0	0.000117	0.001488
в	0.006383	0.000223	0.000359	0.000231	0.000708	0.007617	5.93e-05	0.000847	0.003496	0	0.000899	0.001052
г	0.00113	6.82e-05	7.27e-05	1.19e-05	0.001706	0.000208	7.42e-06	2.67e-05	0.000785	0	0.000163	0.001635
д	0.005073	8.31e-05	0.001268	6.53e-05	8.01e-05	0.005346	0.001335	0.000187	0.00271	0	0.000313	0.001026
е	0.000138	0.002366	0.003658	0.00432	0.004919	0.002111	0.001188	0.002	0.00119	0.002548	0.002124	0.006594
ж	0.001669	2.52e-05	8.90e-06	1.19e-05	0.000961	0.00406	5.93e-06	1.48e-06	0.002047	0	0.000132	1.34e-05
з	0.006875	0.000237	0.00113	0.000516	0.001096	0.000289	6.97e-05	0.000101	0.000654	0	0.000362	0.000322
и	0.000221	0.001698	0.006659	0.000972	0.002796	0.003203	0.000452	0.002836	0.001807	0.001078	0.003708	0.006935
й	9.05e-05	0.000415	0.000745	0.000208	0.000909	0.000169	0.000169	0.000211	0.000617	0	0.001167	0.000546
к	0.008718	0.000331	0.000544	8.75e-05	0.0003	0.00057	0.000159	9.20e-05	0.003139	0	0.000399	0.001193
л	0.006923	0.00046	0.000762	0.000478	0.00055	0.005534	0.000458	0.000223	0.010246	0	0.000863	0.000279
м	0.003103	0.000918	0.000862	0.00026	0.000501	0.004687	0.000141	0.000203	0.006089	0	0.000666	0.000323
н	0.010339	0.000255	0.000494	0.000263	0.000335	0.010271	2.82e-05	0.000196	0.009204	0	0.000699	0.000312
о	0.000117	0.005836	0.009133	0.00596	0.006396	0.002866	0.002502	0.002534	0.001934	0.004757	0.003575	0.008467
п	0.001221	1.48e-06	1.48e-06	1.48e-06	1.48e-06	0.002178	1.48e-06	1.48e-06	0.000927	0	0.000134	0.000994
р	0.008326	0.000957	0.00049	0.000313	0.000616	0.006732	0.000466	0.000132	0.005576	0	0.000398	0.000418
с	0.002185	0.000383	0.002253	0.000129	0.00049	0.003361	0.000141	0.000111	0.002428	0	0.004618	0.004231
т	0.00638	0.0003	0.003771	8.90e-05	0.000383	0.008203	5.34e-05	0.000157	0.004965	0	0.00111	0.001124
у	0.000157	0.000813	0.001556	0.001541	0.002815	0.000524	0.001439	0.000473	0.000848	0.000141	0.001604	0.002255
ф	0.000113	0	1.34e-05	0	4.45e-06	0.000558	0	0	0.000199	0	0	1.19e-05
х	0.000586	0.000179	0.000521	6.23e-05	0.000249	9.64e-05	3.71e-05	8.60e-05	0.0005	0	0.000203	0.0003
ц	0.000681	8.90e-06	8.16e-05	2.97e-06	1.48e-05	0.000897	1.48e-06	7.42e-06	0.000138	0	5.34e-05	2.97e-06
ч	0.002531	1.48e-06	1.19e-05	1.48e-06	2.97e-06	0.003732	0	2.97e-06	0.002394	0	0.000282	4.15e-05
ш	0.001154	5.93e-06	3.26e-05	1.48e-06	0	0.002568	0	0	0.002169	0	0.000798	0.000656

Біграми з пробілами без перекриття (with_sp_bi_nooverlap).

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф
а	0.016076	0.000741	0.002912	0.000796	0.001888	0.000962	0.000846	0.003815	0.000129	0.001348	0.004524	0.000865	0.003183	0.002616	9.95e-06	0.000609	0.002221	0.003457	0.003952	0.000418	8.21e-05
б	0.000231	0.001236	7.46e-06	5.47e-05	2.49e-06	2.74e-05	0.002124	2.49e-06	0	0.000888	0	8.21e-05	0.001194	7.71e-05	0.000226	0.002305	0	0.00141	0.000149	2.49e-06	0.001176
в	0.005302	0.005116	7.46e-06	3.48e-05	0	0.000256	0.006277	0	0.000597	0.002604	0	0.000196	0.000749	0.000107	0.000865	0.006011	0.000291	0.000933	0.001865	0.000244	0.000798
г	0.000761	0.000953	0	7.46e-06	0	0.001363	0.000142	0	0.000589	0	9.45e-05	0.001338	7.46e-06	0.000321	0.007665	0	0.001042	1.99e-05	1.99e-05	0.00047	0
д	0.001179	0.004223	3.73e-05	0.00098	2.98e-05	1.99e-05	0.004452	0.000972	7.21e-05	0.002141	0	0.000201	0.000898	7.96e-05	0.001746	0.004123	5.47e-05	0.000933	0.000353	0.000241	0.001333
е	0.016491	1.74e-05	0.001107	0.001731	0.00331	0.003474	0.0014	0.000821	0.001074	9.45e-05	0.002149	0.001002	0.00521	0.003616	0.007503	0.000144	0.000766	0.007996	0.00432	0.004646	0.000159
ж	0.000109	0.001522	7.46e-06	0	7.46e-06	0.000848	0.003504	4.97e-06	0	0.00183	0	0.000117	7.46e-06	2.49e-06	0.000826	3.23e-05	0	4.97e-06	9.95e-06	0	0.000167
з	0.001455	0.005865	0.000157	0.000828	0.000408	0.000791	0.000239	3.73e-05	2.98e-05	0.000505	0	0.000162	0.000251	0.000281	0.001864	0.00044	0	0.000338	0	1.74e-05	0.000425
и	0.018324	4.97e-05	0.000445	0.003925	0.000458	0.001656	0.002241	0.000226	0.001811	0.000448	0.000915	0.002151	0.005648	0.002718	0.003037	0.000109	0.000167	0.000629	0.0047	0.003974	7.46e-06
й	0.006108	0	0	4.97e-06	4.97e-06	0.000196	9.95e-06	0	0	0	0	0.000497	0.000333	3.98e-05	0.000321	4.97e-06	5.47e-05	2.49e-06	0.000286	0.00055	0
к	0.004392	0.007277	0	9.95e-05	0	0	0.000395	4.48e-05	2.49e-06	0.002355	0	9.95e-06	0.000958	0	0.000634	0.008043	0	0.001768	0.000632	0.000532	0.001539
л	0.007929	0.005633	2.24e-05	5.22e-05	0.000157	5.97e-05	0.004158	0.000373	1.49e-05	0.008108	0	0.00042	0.000167	1.99e-05	0.000311	0.005797	9.70e-05	0	0.001651	0.000142	0.001681
м	0.007904	0.002477	0.000527	0	2.98e-05	2.49e-06	0.003837	0	2.49e-06	0.004479	0	0.000102	0.000179	3.48e-05	0.0015	0.003738	8.21e-05	9.70e-05	0.000109	4.97e-06	0.002161
н	0.004101	0.008645	4.97e-06	1.24e-05	0.000139	0.000172	0.008481	2.49e-06	2.24e-05	0.007456	0	0.00043	0.000236	0	0.002216	0.008846	0	4.97e-05	0.000492	0.000816	0.002984
о	0.019399	0	0.004007	0.005991	0.004738	0.004521	0.00189	0.001853	0.001515	0.000768	0.003959	0.002305	0.0067	0.005598	0.005452	0.000147	0.001194	0.004855	0.006183	0.007033	0.000311
п	3.73e-05	0.001012	0	0	0	0.001808	0	0	0.000754	0	0.000114	0.000853	0	0.000144	0.008881	1.24e-05	0.006703	2.49e-06	9.70e-05	0.000781	0
р	0.002664	0.007011	0.000154	0.000274	0.000119	0.000376	0.005544	0.000366	7.71e-05	0.004653	0	0.000229	0.000306	0.000204	0.00098	0.006828	0.000104	8.46e-05	0.000139	0.00048	0.002358
с	0.004176	0.001778	9.70e-05	0.00147	2.49e-05	0.000311	0.002674	4.73e-05	7.46e-06	0.001808	0	0.003639	0.003574	0.000863	0.001052	0.002283	0.001726	0.000122	0.000838	0.010993	0.000778
т	0.00522	0.005208	2.74e-05	0.002869	2.49e-06	0.000122	0.006663	0	0	0.003885	0	0.000709	0.000808	2.49e-05	0.001079	0.012452	0.000142	0.002291	0.001191	9.70e-05	0.001263
у	0.006772	3.48e-05	0.000495	0.000721	0.001112	0.002069	0.000293	0.001059	0.000239	0.000152	0.000112	0.00098	0.00186	0.001054	0.000505	1.24e-05	0.000676	0.000269	0.001161	0.001559	0
ф	4.73e-05	0.000109	0	0	0	0.00041	0	0	0.000164	0	0	1.24e-05	2.49e-06	0	0.000102	0	2.98e-05	7.46e-06	2.49e-06	8.95e-05	1.74e-05
х	0.003106	0.000465	0	0.000231	0	0	3.48e-05	0	0.000216	0	0	0.000154	5.47e-05	0.000199	0.002024	0	0.000112	3.23e-05	0	0.000129	2.49e-06
ц	0.000323	0.000567	0	4.23e-05	0	0.000711	0	0	8.95e-05	0	9.95e-06	0	0	0	0.000214	0	0	0	0	0.000112	0
ч	8.95e-05	0.002084	0	9.95e-06	0	0.003116	0	0.000197	0	0.00197	0	0.000249	3.23e-05	2.49e-06	0.000592	6.71e-05	0	3.23e-05	0	0.002999	0.000535
ш	5.22e-05	0.000982	0	2.24e-05	0	0.002169	0	0.001763	0	0.001763	0	0.000629	0.000522	2.98e-05	0.000199	0.000147	1.24e-05	0	0	7.21e-05	0.000284

	а	б	в	г	д	е	ж	з	и	й	к	л
	0.001194	0.007819	0.014039	0.003216	0.007618	0.004626	0.001512	0.004502	0.0091	0	0.008687	0.002609
а	0.016076	0.000741	0.002912	0.000798	0.001888	0.000962	0.000846	0.003815	0.000129	0.001348	0.004524	0.008085
б	0.000231	0.001236	7.46e-06	5.47e-05	2.49e-06	2.74e-05	0.002124	2.49e-06	0	0.000888	8.21e-05	0.001194
в	0.005302	0.005116	7.46e-06	3.48e-05	0	0.000256	0.006277	0	0.000597	0.002604	0	0.000196
г	0.000761	0.000953	0	7.46e-06	0	0.001363	0.000142	0	0	0.000589	0	9.45e-05
д	0.001179	0.004223	3.73e-05	0.00098	2.98e-05	1.99e-05	0.004452	0.000972	7.21e-05	0.002141	0	0.000201
е	0.016491	1.74e-05	0.001107	0.001731	0.00331	0.003474	0.0014	0.000821	0.001074	9.45e-05	0.002149	0.001002
ж	0.000109	0.001522	7.46e-06	0	7.46e-06	0.000848	0.003504	4.97e-06	0	0.00183	0	7.46e-06
з	0.001455	0.005665	0.000157	0.000828	0.000408	0.000791	0.000239	3.73e-05	2.98e-05	0.000505	0	0.000162
и	0.018324	4.97e-05	0.000445	0.003925	0.000458	0.001656	0.002241	0.000226	0.001811	0.000448	0.000915	0.002151
й	0.008108	0	0	4.97e-06	4.97e-06	0.000196	9.95e-06	0	0	0	0	0.000497
к	0.004392	0.007277	0	9.95e-05	0	0	0.000395	4.48e-05	2.49e-06	0.002355	0	9.95e-06
л	0.007929	0.005633	2.24e-05	5.22e-05	0.000157	5.97e-05	0.004158	0.000373	1.49e-05	0.008108	0	0.00042
м	0.007904	0.002477	0.000527	0	2.98e-05	2.49e-06	0.003837	0	2.49e-06	0.004479	0	0.000102
н	0.004101	0.008645	4.97e-06	1.24e-05	0.000139	0.000172	0.008481	2.49e-06	2.24e-05	0.007456	0	0.00043
о	0.019399	0	0.004007	0.005991	0.004738	0.004521	0.00189	0.001853	0.001515	0.000768	0.003959	0.002305
п	3.73e-05	0.001012	0	0	0	0	0.001808	0	0	0.000754	0	0.000114
р	0.002664	0.007011	0.000154	0.000274	0.000119	0.000376	0.005544	0.000366	7.71e-05	0.004653	0	0.000229
с	0.004176	0.001778	9.70e-05	0.00147	2.49e-05	0.000311	0.002674	4.73e-05	7.46e-06	0.001808	0	0.003639
т	0.00522	0.005208	2.74e-05	0.002669	2.49e-06	0.000122	0.006663	0	0	0.003885	0	0.000709
у	0.006772	3.48e-05	0.000495	0.000721	0.001112	0.002069	0.000293	0.001059	0.000239	0.000152	0.000112	0.00098
ф	4.73e-05	0.000109	0	0	0	0	0.00041	0	0	0.000164	0	1.24e-05
х	0.003106	0.000465	0	0.000231	0	0	3.48e-05	0	0	0.000216	0	0.000154
ц	0.000323	0.000567	0	4.23e-05	0	0	0.000711	0	0	8.95e-05	0	9.95e-06
ч	8.95e-05	0.002084	0	9.95e-06	0	0	0.003116	0	0	0.00197	0	0.000249
ш	5.22e-05	0.000982	0	2.24e-05	0	0	0.002169	0	0	0.001763	0	0.000629

Біграми без пробілів та без перекриття (no_sp_bi_nooverlap).

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х
а	0.000136	0.001605	0.004925	0.001264	0.003148	0.001688	0.00132	0.005052	0.001296	0.001644	0.006435	0.009654	0.004643	0.005195	0.001626	0.002471	0.003174	0.006037	0.005477	0.000946	0.000184	0.001136
б	0.001451	8.90e-06	9.20e-05	0	2.97e-05	0.002557	1.19e-05	8.90e-06	0.001148	0	0.000104	0.00146	8.01e-05	0.000258	0.002771	2.97e-06	0.001706	0.000184	1.48e-05	0.001409	0	7.12e-05
в	0.006328	0.000225	0.000365	0.000193	0.000679	0.007619	5.04e-05	0.000872	0.003344	0	0.000902	0.001041	0.000291	0.001516	0.007536	0.000961	0.001362	0.002896	0.00086	0.001065	2.97e-06	7.12e-05
г	0.001154	8.01e-05	5.93e-05	1.19e-05	0.00167	0.000208	1.19e-05	2.67e-05	0.00084	0	0.000148	0.001605	5.64e-05	0.000436	0.00925	9.79e-05	0.001127	0.000119	3.86e-05	0.000638	0	2.97e-06
д	0.005159	6.23e-05	0.001243	6.82e-05	8.31e-05	0.005497	0.00135	0.000199	0.002697	0	0.000329	0.001092	0.000166	0.002409	0.005011	0.000172	0.00119	0.00051	0.000383	0.001682	2.97e-06	9.79e-05
е	0.000145	0.002367	0.003625	0.004361	0.004925	0.002041	0.001139	0.002029	0.001234	0.002519	0.002151	0.006411	0.005595	0.010577	0.001715	0.003005	0.009873	0.007277	0.00656	0.000902	7.12e-05	0.000697
ж	0.001682	2.67e-05	5.93e-06	1.48e-05	0.000917	0.00405	2.97e-06	0	0.002044	0	0.000131	1.48e-05	2.97e-06	0.001107	5.34e-05	1.78e-05	2.97e-06	2.67e-05	8.90e-06	0.000148	0	5.93e-06
з	0.00688	0.000267	0.001104	0.000534	0.001086	0.000258	7.12e-05	0.000119	0.000656	0	0.000406	0.000317	0.000401	0.00205	0.00062	0.000178	0.00049	0.000139	7.71e-05	0.000475	2.97e-06	0
и	0.000199	0.001623	0.006953	0.001008	0.002762	0.00324	0.000424	0.002807	0.001836	0.001032	0.003684	0.008986	0.004234	0.005373	0.00194	0.002531	0.001145	0.007992	0.005675	0.000679	3.86e-05	0.002216
й	9.49e-05	0.000436	0.000807	0.000255	0.000917	0.000175	0.000163	0.000214	0.000555	0	0.001169	0.000531	0.000771	0.001035	0.000679	0.000958	0.000341	0.001448	0.001	0.000223	6.82e-05	5.93e-05
к	0.008663	0.000329	0.00059	9.49e-05	0.000282	0.000608	0.000184	0.00011	0.003005	0	0.000436	0.001166	0.000202	0.001261	0.009986	0.000401	0.002255	0.001296	0.000994	0.002056	2.08e-05	3.86e-05
л	0.006924	0.00049	0.000822	0.000469	0.000531	0.005515	0.000442	0.000217	0.010535	0	0.000851	0.000255	0.000914	0.00116	0.007948	0.000813	0.000205	0.00318	0.00043	0.002207	0.000196	4.15e-05
м	0.003029	0.000937	0.00086	0.000291	0.000501	0.004643	0.000128	0.000193	0.006091	0	0.000673	0.000291	0.000392	0.002456	0.005269	0.001095	0.000338	0.001237	0.000312	0.002913	8.60e-05	9.20e-05
н	0.01039	0.000258	0.000507	0.000255	0.000338	0.010274	3.56e-05	0.000211	0.009191	0	0.000682	0.000323	0.000157	0.003166	0.010864	0.000617	0.000187	0.001098	0.001175	0.003797	1.78e-05	8.31e-05
о	0.000113	0.005865	0.000917	0.005954	0.006233	0.002774	0.002501	0.002563	0.001887	0.004732	0.00364	0.008369	0.008004	0.008817	0.002195	0.003803	0.00645	0.009707	0.009428	0.000108	0.000142	0.00078
п	0.001178	2.97e-06	2.97e-06	0	0	0.002166	2.97e-06	0	0.000881	0	0.000131	0.000982	0	0.000157	0.010401	1.78e-05	0.007817	8.90e-06	0.000122	0.000935	0	8.1e-06
р	0.008242	0.000982	0.000424	0.000344	0.000611	0.006702	0.000484	0.000113	0.00558	0	0.000401	0.000439	0.000297	0.001359	0.008141	0.000433	0.000131	0.000534	0.000668	0.002907	8.90e-05	0.000142
с	0.002204	0.000421	0.002287	0.000113	0.00043	0.003438	0.000148	0.000134	0.002501	0	0.004507	0.00415	0.001445	0.001741	0.003014	0.00259	0.000249	0.001296	0.013778	0.000997	2.97e-05	0.000326
т	0.006468	0.000285	0.003854	8.60e-05	0.000406	0.00823	6.23e-05	0.000166	0.004946	0	0.001133	0.001124	0.000314	0.001911	0.015166	0.000694	0.002675	0.001949	0.000406	0.001715	3.86e-05	0.000104
у	0.000154	0.000613	0.001563	0.001501	0.002839	0.000587	0.001391	0.000463	0.000816	0.000145	0.001578	0.002302	0.001736	0.001584	0.00059	0.001522	0.000492	0.00216	0.002166	0.000151	5.93e-05	0.00059
ф	0.000113	0	8.90e-06	0	8.90e-06	0.000576	0	0	0.000193	0	0	8.90e-06	2.97e-06	2.97e-06	0.000148	2.97e-06	2.97e-05	2.67e-05	2.97e-06	0.000125	1.78e-05	0
х	0.00062	0.00016	0.000549	7.12e-05	0.000228	9.49e-05	5.34e-05	9.49e-05	0.000498	0	0.000184	0.000303	0.00022	0.000448	0.002418	0.000406	0.000243	0.00046	0.000148	0.000276	3.56e-05	2.97e-05
ц	0.0007	5.93e-06	9.49e-05	2.97e-06	1.19e-05	0.000881	2.97e-06	8.90e-06	0.000134	0	5.64e-05	2.97e-06	2.37e-05	3.86e-05	0.000326	3.56e-05	2.97e-06	1.78e-05	8.90e-06	0.000154	2.97e-06	5.93e-06
ч	0.00254	2.97e-06	1.48e-05	2.97e-06	5.93e-06	0.003682	0	2.97e-06	0.002442	0	0.000312	3.56e-05	8.90e-06	0.000685	9.20e-05	5.93e-06	4.75e-05	1.19e-05	0.003673	0.000817	0	0
ш	0.001142	8.90e-06	3.26e-05	0	0	0.002587	0	0	0.0021	0	0.00081	0.000653	6.23e-05	0.000217	0.000181	2.97e-05	1.19e-05	2.97e-06	6.23e-05	0.000312	0	2.97e-06
щ	0.00051	2.97e-06	2.97e-06	0	0	0.001391	0	0	0.001118	0	0	0	0	7.42e-05	1.48e-05	0	2.97e-06	2.97e-06	2.97e-06	8.01e-05	0	2.97e-06

	а	б	в	г	д	е	ж	з	и	й	к
а	0.000136	0.001605	0.004925	0.001264	0.003148	0.001688	0.00132	0.005052	0.001296	0.001644	0.006435
б	0.001451	8.90e-06	9.20e-05	0	2.97e-05	0.002557	1.19e-05	8.90e-06	0.001148	0	0.000104
в	0.006328	0.000225	0.000365	0.000193	0.000679	0.007619	5.04e-05	0.000872	0.003344	0	0.000902
г	0.001154	8.01e-05	5.93e-05	1.19e-05	0.00167	0.000208	1.19e-05	2.67e-05	0.00084	0	0.000148
д	0.005159	6.23e-05	0.001243	6.82e-05	8.31e-05	0.005497	0.00135	0.000199	0.002697	0	0.000329
е	0.000145	0.002367	0.003625	0.004361	0.004925	0.002041	0.001139	0.002029	0.001234	0.002519	0.002151
ж	0.001682	2.67e-05	5.93e-06	1.48e-05	0.000917	0.00405	2.97e-06	0	0.002044	0	0.000131
з	0.00688	0.000267	0.001104	0.000534	0.001086	0.000258	7.12e-05	0.000119	0.000656	0	0.000406
и	0.000199	0.001623	0.006693	0.001006	0.002762	0.00324	0.000424	0.002807	0.001836	0.001032	0.003664
й	9.49e-05	0.000436	0.000807	0.000255	0.000917	0.000175	0.000163	0.000214	0.000555	0	0.001169
к	0.008663	0.000329	0.00059	9.49e-05	0.000282	0.000608	0.000184	0.00011	0.003005	0	0.000436
л	0.006924	0.00049	0.000822	0.000469	0.000531	0.005515	0.000442	0.000217	0.010535	0	0.000851
м	0.003029	0.000937	0.00086	0.000291	0.000501	0.004643	0.000128	0.000193	0.006091	0	0.000673
н	0.01039	0.000258	0.000507	0.000255	0.000338	0.010274	3.56e-05	0.000211	0.009191	0	0.000682
о	0.000113	0.005865	0.00917	0.005954	0.006233	0.002774	0.002501	0.002563	0.001887	0.004732	0.00364
п	0.001178	2.97e-06	2.97e-06	0	0	0.002166	2.97e-06	0	0.000881	0	0.000131
р	0.008242	0.000982	0.000424	0.000344	0.000611	0.006702	0.000484	0.000113	0.00558	0	0.000401
с	0.002204	0.000421	0.002287	0.000113	0.00043	0.003438	0.000148	0.000134	0.002501	0	0.004507
т	0.006468	0.000285	0.003854	8.60e-05	0.000406	0.00823	6.23e-05	0.000166	0.004946	0	0.001133
у	0.000154	0.000813	0.001563	0.001501	0.002839	0.000587	0.001391	0.000463	0.000816	0.000145	0.001578
ф	0.000113	0	8.90e-06	0	8.90e-06	0.000576	0	0	0.000193	0	0
х	0.00062	0.00016	0.000549	7.12e-05	0.000228	9.49e-05	5.34e-05	9.49e-05	0.000498	0	0.000184
ц	0.0007	5.93e-06	9.49e-05	2.97e-06	1.19e-05	0.000881	2.97e-06	8.90e-06	0.000134	0	5.64e-05
ч	0.00254	2.97e-06	1.48e-05	2.97e-06	5.93e-06	0.003682	0	2.97e-06	0.002442	0	0.000312
ш	0.001142	8.90e-06	3.26e-05	0	0	0.002587	0	0	0.0021	0	0.00081

Таблиці:

with_sp_bi_overlap – з пробілами та перекриттям.

no_sp_bi_overlap – без пробілів та з перекриттям.

with_sp_bi_nooverlap – з пробілами без перекриття.

no_sp_bi_nooverlap – без пробілів та без перекриття.

Оцінка значень ентропії та надлишковості

H_1 – монограми, H_2 – біграми.

	Ентропія, H (біт/симв.)	Надлишковість, R
H_1 (з пробілами)	4.3806269868	0.123875
H_2 (з пробілами, overlapped)	3.9743496766	0.205130
H_2 (з пробілами, non-overlapped)	3.9740673622	0.205187
H_1 (без пробілів)	4.4641246350	0.098921
H_2 (без пробілів, overlapped)	4.1440491334	0.163527
H_2 (без пробілів, non-overlapped)	4.1431752005	0.163704

Було досліджено чотири варіанти біграм:

1. З пробілом і з перекриттям
2. З пробілом без перекриття
3. Без пробілу з перекриттям
4. Без пробілу без перекриття

Порівняння результатів:

[illegible]

$H^{(20)}$:

Лабораторная работа №1

Произвольная часть текста:
не_считался_похвальным_качеством_разного_мнения_держались_люди_и_по_тому_во

Использованные буквы:

Порядок n-граммы:
5 символов
10 символов
15 символов
20 символов
25 символов
30 символов
35 символов
40 символов
45 символов
50 символов

Введенный символ: н
Символ по счету: 1
Номер эксперимента: 51
Поле ввода символов:
н
Продолжить Другой

Неравенство для энтропии:
 $1,9167458982552 < H < 2,59530861634623$
Двоичная таблица угаданных символов:
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
00000001000000000000000000000000
10000000000000000000000000000000
~~~~~

Вероятности:  
q[1] = 0,4705882  
q[2] = 0,1764705  
q[3] = 0,0980392  
q[4] = 0  
q[5] = 0  
q[6] = 0,0588235  
q[7] = 0  
q[8] = 0,0392156  
q[9] = 0  
q[10] = 0  
q[11] = 0,019607  
q[12] = 0,019607  
q[13] = 0  
q[14] = 0,019607  
q[15] = 0  
q[16] = 0  
q[17] = 0  
q[18] = 0  
q[19] = 0  
q[20] = 0  
q[21] = 0,019607  
q[22] = 0,019607  
q[23] = 0  
q[24] = 0  
q[25] = 0  
q[26] = 0  
q[27] = 0,019607  
q[28] = 0  
q[29] = 0  
q[30] = 0,019607  
q[31] = 0  
q[32] = 0,019607

Строка состояния:  
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

$H^{(30)}$ :

Лабораторная работа №1

Произвольная часть текста:  
мы\_ставим\_его\_исключительно\_в\_заслугу\_себе\_итак\_я\_хочу\_обратить\_ваше\_вниман

Использованные буквы:

Порядок n-граммы:  
5 символов  
10 символов  
15 символов  
20 символов  
25 символов  
30 символов  
35 символов  
40 символов  
45 символов  
50 символов

Введенный символ: \_ (пробел)  
Символ по счету: 1  
Номер эксперимента: 51  
Поле ввода символов:  
Продолжить Другой

Неравенство для энтропии:  
 $1,48100572462754 < H < 2,13389581375916$   
Двоичная таблица угаданных символов:  
10000000000000000000000000000000  
10000000000000000000000000000000  
10000000000000000000000000000000  
10000000000000000000000000000000  
10000000000000000000000000000000  
~~~~~

Вероятности:
q[1] = 0,6470588
q[2] = 0,0784313
q[3] = 0,0392156
q[4] = 0,0392156
q[5] = 0
q[6] = 0,0196078
q[7] = 0
q[8] = 0
q[9] = 0,0196078
q[10] = 0
q[11] = 0
q[12] = 0
q[13] = 0,019607
q[14] = 0
q[15] = 0
q[16] = 0,019607
q[17] = 0
q[18] = 0,019607
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0
q[24] = 0
q[25] = 0,019607
q[26] = 0
q[27] = 0,039215
q[28] = 0
q[29] = 0,019607
q[30] = 0
q[31] = 0,019607
q[32] = 0

Строка состояния:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

Для 10 символів

2.31062508101796 < H < 3.11332989948939

Для 20 символів

1.9167458982552 < H < 2.59530861634623

Для 35 символів

1.48100572462754 < H < 2.13339581375916

Зі збільшенням довжини контексту ентропія зменшується, оскільки враховується більше інформації про попередні символи, що робить джерело більш передбачуваним.

$$H_1 = - \sum_{i=1}^m p_i \log_2 p_i$$

$$H_2 = - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m p_{ij} \log_2 p_{ij}$$

- H_1 — ентропія монограм (символів);
- H_2 — ентропія біграм (пар символів) у перерахунку на один символ;
- p_i — ймовірність появи символу i ;
- p_{ij} — ймовірність появи біграми ij ;
- m — кількість символів в алфавіті.

Надлишковість обчислюємо за формулою:

$$R = 1 - \frac{H}{\log_2 m}$$

де m — розмір алфавіту (у нашому випадку близько 32 символів, тобто $\log_2 m \approx 5$).

Надлишковість R для різних кількостей символів:

Для 10 символів:

$$R_{\min} = 0.377334$$

$$R_{\max} = 0.537875$$

Для 20 символів:

$$R_{\min} = 0.480938$$

$$R_{\max} = 0.616651$$

Для 35 символів:

$$R_{\min} = 0.573321$$

$$R_{\max} = 0.703799$$

Розрахунки показали, що надлишковість зростає зі збільшенням довжини контексту:

- для 10 символів — приблизно 40–55 %,
- для 20 символів — 48–62 %,
- для 35 символів — 57–70 %.

Зі зростанням довжини контексту ентропія поступово зменшується, а надлишковість — зростає.

Це свідчить, що у довших послідовностях міститься більше повторюваної та структурованої інформації, завдяки чому текст стає більш передбачуваним і стійким до втрат або спотворень символів.

Висновки:

У ході роботи ми ознайомилися з поняттями ентропії та надлишковості й на практиці переконалися, як вони проявляються в аналізі тексту. Ми побачили, що ентропія відображає ступінь невизначеності (або кількість нової інформації), а надлишковість — рівень повторюваності та структурованості мови.

Під час експериментів спостерігалось, що при врахуванні більшої кількості символів контекст стає сильнішим: ентропія зменшується, а надлишковість збільшується. Це показує, що мова має певні правила, закономірності й обмеження, завдяки яким повідомлення залишаються зрозумілими навіть при втраті частини інформації.

Отже, робота дала змогу закріпити теоретичні знання про ентропію, навчитися обчислювати її на практиці та побачити, як за допомогою формул можна оцінити інформаційні властивості тексту і ступінь його надлишковості.