

МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМ.ІГОРЯ СІКОРСЬКОГО»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра інформаційної безпеки

Комп'ютерний практикум №1

Виконали:

Студенти 3 курсу

ФБ-01 Літвінчук Софія
та
ФБ-02 Косарик Дарія

Тема: Експериментальна оцінка ентропії на символ джерела відкритого тексту

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.

Постановка задачі:

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.

Файл з текстом: text.txt

Файл з кодом: main.py

Під час роботи було створено такі таблиці, які містять відповідну інформацію:

- fr.bi.cross.off.xlsx - частоти біграм, букви яких перетинаються у тексті без пробілів
- fr.bi.cross.xlsx - частоти біграм, букви яких перетинаються у тексті з пробілами
- fr.bi.notcross.off.xlsx - частоти біграм, букви яких не перетинаються у тексті без пробілів
- fr.bi.notcross.xlsx - частоти біграм, букви яких не перетинаються у тексті з пробілами
- fr.letters.off.xlsx - частоти букв у тексті без пробілів
- fr.letters.xlsx - частоти букв у тексті з пробілами

Хід роботи:

0. Підготовка тексту.

- залишили лише слова російською мовою, змінили регістр, видалили зайві пробіли.

1. Таблиці частот

Для тексту без пробілів

➤ для букв

	о	е	а	н	и	т	с	л	в	р	к	д	м
frequency	0,114723	0,086829	0,083208	0,06909	0,066403	0,060297	0,053609	0,050233	0,046491	0,039503	0,034877	0,029652	0,028791

у	п	я	ь	ы	г	б	ч	з	ж	й	ш	х	ю	э	щ	ц	ф	ъ	ё
0,027077	0,024094	0,021563	0,020055	0,018578	0,018074	0,017716	0,016918	0,016649	0,01133	0,01074	0,008585	0,007914	0,006384	0,003613	0,002832	0,002716	0,001124	0,000307	2,43E-05

➤ для біграм, букви яких перетинаються

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
а	0,000494	0,001589	0,005524	0,001094	0,003632	0,002052	0,001529	0,005575	0,001649	0,000683	0,006349	0,012241	0,003652	0,007701	0,001963
б	0,001086	2,86E-06	9,43E-05	2E-05	4E-05	0,0024	8,57E-06	2,86E-06	0,000974	0	0,000311	0,000994	9,43E-05	0,000377	0,002446
в	0,006681	0,000166	0,00038	0,00032	0,00072	0,005509	4,86E-05	0,000694	0,005401	0	0,000763	0,001117	0,000331	0,001777	0,008792
г	0,001009	5,14E-05	5,43E-05	5,71E-06	0,001409	0,000354	2E-05	2E-05	0,00094	0	0,000174	0,001892	1,43E-05	0,000263	0,009915
д	0,005063	9,72E-05	0,000974	2E-05	0,0001	0,005372	2,29E-05	6E-05	0,002426	0	0,000417	0,000983	0,000134	0,001935	0,004386
е	0,000349	0,002552	0,004718	0,004981	0,003903	0,003	0,001157	0,002317	0,001537	0,003015	0,003135	0,007335	0,005995	0,010393	0,00186
ж	0,001466	7,14E-05	3,14E-05	8,57E-06	0,000949	0,004758	8,57E-06	8,57E-06	0,001809	0	0,000129	8,57E-06	1,71E-05	0,001149	6,86E-05
з	0,006146	0,000197	0,001069	0,000597	0,00096	0,000374	0,000174	9,43E-05	0,000383	0	0,000246	0,000209	0,000406	0,00244	0,000549
и	0,000363	0,001329	0,005209	0,000869	0,00276	0,003426	0,000612	0,002686	0,001966	0,001637	0,003103	0,005772	0,004232	0,007138	0,002003
й	0,000474	0,00034	0,00068	0,000237	0,0008	0,000126	0,000189	0,000211	0,000752	0	0,000617	0,000266	0,000406	0,000969	0,000569
к	0,00945	0,000409	0,00056	8,57E-05	0,000237	0,000697	0,000246	8,86E-05	0,003789	0	0,000343	0,000509	0,000237	0,001446	0,009844
л	0,009387	0,000243	0,000797	0,000243	0,000274	0,006661	0,0005	0,000143	0,006569	0	0,000846	0,000531	0,000154	0,00106	0,008104
м	0,003026	0,00036	0,000714	0,000269	0,0004	0,003563	0,000137	0,000211	0,003715	0	0,000577	0,000343	0,000254	0,002135	0,004343
н	0,01351	0,00034	0,000712	0,000226	0,001046	0,011773	6E-05	0,000206	0,00913	0	0,000469	0,000131	0,000111	0,004089	0,011913
о	0,000283	0,005258	0,010973	0,005935	0,006084	0,00366	0,002963	0,001829	0,002369	0,003506	0,003115	0,007047	0,006681	0,012741	0,00292
п	0,001443	0	2,86E-06	0	2,86E-06	0,002386	2,86E-06	2,86E-06	0,000886	0	8,86E-05	0,000557	2,86E-06	0,000114	0,00967
р	0,007958	0,000334	0,00048	0,000426	0,000466	0,006244	0,000351	3,43E-05	0,005469	0	0,000743	6,86E-05	0,000186	0,000852	0,008292
с	0,002186	0,000177	0,002477	0,000131	0,000612	0,005406	0,000249	9,14E-05	0,001686	0	0,006201	0,002972	0,001034	0,001606	0,002972
т	0,006101	0,000311	0,00354	7,14E-05	0,000346	0,006095	8,86E-05	0,0001	0,005098	0	0,000852	0,000417	0,000183	0,001743	0,017731
у	0,000206	0,00078	0,001983	0,001292	0,002323	0,000426	0,002195	0,000549	0,000814	0,000146	0,001363	0,001686	0,00154	0,0012	0,000531

➤ для біграм, букви яких не перетинаються

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
а	0,000494	0,001589	0,005524	0,001094	0,003632	0,002052	0,001529	0,005575	0,001649	0,000683	0,006349	0,012241	0,003652	0,007701	0,001963
б	0,001086	2,86E-06	9,43E-05	2E-05	4E-05	0,0024	8,57E-06	2,86E-06	0,000974	0	0,000311	0,000994	9,43E-05	0,000377	0,002446
в	0,006681	0,000166	0,00038	0,00032	0,00072	0,005509	4,86E-05	0,000694	0,005401	0	0,000763	0,001117	0,000331	0,001777	0,008792
г	0,001009	5,14E-05	5,43E-05	5,71E-06	0,001409	0,000354	2E-05	2E-05	0,00094	0	0,000174	0,001892	1,43E-05	0,000263	0,009915
д	0,005063	9,72E-05	0,000974	2E-05	0,0001	0,005372	2,29E-05	6E-05	0,002426	0	0,000417	0,000983	0,000134	0,001935	0,004386
е	0,000349	0,002552	0,004718	0,004981	0,003903	0,003	0,001157	0,002317	0,001537	0,003015	0,003135	0,007335	0,005995	0,010393	0,00186
ж	0,001466	7,14E-05	3,14E-05	8,57E-06	0,000949	0,004758	8,57E-06	8,57E-06	0,001809	0	0,000129	8,57E-06	1,71E-05	0,001149	6,86E-05
з	0,006146	0,000197	0,001069	0,000597	0,00096	0,000374	0,000174	9,43E-05	0,000383	0	0,000246	0,000209	0,000406	0,00244	0,000549
и	0,000363	0,001329	0,005209	0,000869	0,00276	0,003426	0,000612	0,002686	0,001966	0,001637	0,003103	0,005772	0,004232	0,007138	0,002003
й	0,000474	0,00034	0,00068	0,000237	0,0008	0,000126	0,000189	0,000211	0,000752	0	0,000617	0,000266	0,000406	0,000969	0,000569
к	0,00945	0,000409	0,00056	8,57E-05	0,000237	0,000697	0,000246	8,86E-05	0,003789	0	0,000343	0,000509	0,000237	0,001446	0,009844
л	0,009387	0,000243	0,000797	0,000243	0,000274	0,006661	0,0005	0,000143	0,006569	0	0,000846	0,000531	0,000154	0,00106	0,008104
м	0,003026	0,00036	0,000714	0,000269	0,0004	0,003563	0,000137	0,000211	0,003715	0	0,000577	0,000343	0,000254	0,002135	0,004343
н	0,01351	0,00034	0,000712	0,000226	0,001046	0,011773	6E-05	0,000206	0,00913	0	0,000469	0,000131	0,000111	0,004089	0,011913
о	0,000283	0,005258	0,010973	0,005935	0,006084	0,00366	0,002963	0,001829	0,002369	0,003506	0,003115	0,007047	0,006681	0,012741	0,00292
п	0,001443	0	2,86E-06	0	2,86E-06	0,002386	2,86E-06	2,86E-06	0,000886	0	8,86E-05	0,000557	2,86E-06	0,000114	0,00967
р	0,007958	0,000334	0,00048	0,000426	0,000466	0,006244	0,000351	3,43E-05	0,005469	0	0,000743	6,86E-05	0,000186	0,000852	0,008292
с	0,002186	0,000177	0,002477	0,000131	0,000612	0,005406	0,000249	9,14E-05	0,001686	0	0,006201	0,002972	0,001034	0,001606	0,002972
т	0,006101	0,000311	0,00354	7,14E-05	0,000346	0,006095	8,86E-05	0,0001	0,005098	0	0,000852	0,000417	0,000183	0,001743	0,017731
у	0,000206	0,00078	0,001983	0,001292	0,002323	0,000426	0,002195	0,000549	0,000814	0,000146	0,001363	0,001686	0,00154	0,0012	0,000531

Для тексту з пробілами

➤ для букв

	о	е	а	н	и	т	с	л	в	р	к	д	м	у	п	
frequency	0,167356	0,095524	0,072297	0,069283	0,057527	0,05529	0,050206	0,044637	0,041826	0,038711	0,032892	0,02904	0,02469	0,023972	0,022546	0,020062

п	я	ь	ы	г	б	ч	з	ж	й	ш	х	ю	э	щ	ц	ф	ъ	ё
0,020062	0,017954	0,016699	0,015469	0,015049	0,014751	0,014086	0,013863	0,009434	0,008942	0,007148	0,006589	0,005315	0,003009	0,002358	0,002261	0,000936	0,000256	2,02E-05

➤ для біграм, букви яких перетинаються

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
а	1,07E-05	0,000641	0,002993	0,000537	0,002173	0,001004	0,001118	0,004134	0,000128	0,000538	0,004119	0,009709	0,002472	0,00434	7,14E-06
б	0,000897	1,19E-06	6,78E-05	7,14E-06	3,33E-05	0,001939	5,95E-06	0	0,000814	0	0,000251	0,000833	6,9E-05	0,000307	0,001884
в	0,005355	1,31E-05	2,26E-05	3,81E-05	0,000272	0,004215	1,19E-06	0,000538	0,004509	0	0,000156	0,000772	0,00015	0,001194	0,006864
г	0,000808	1,19E-06	1,19E-06	0	0,001161	0,000277	0	0	0,000791	0	0,000136	0,001535	1,19E-06	0,000174	0,008229
д	0,004251	2,14E-05	0,000754	4,76E-06	3,33E-05	0,004461	1,07E-05	2,38E-06	0,001946	0	0,000249	0,000777	7,49E-05	0,001529	0,00369
е	3,93E-05	0,001299	0,002323	0,00362	0,002424	0,002109	0,000719	0,001263	9,52E-05	0,002518	0,001748	0,005701	0,003939	0,007026	0,000184
ж	0,001342	4,76E-05	0	8,33E-06	0,000758	0,003982	5,95E-06	0	0,001476	0	8,45E-05	1,19E-06	2,38E-06	0,000972	4,4E-05
з	0,005227	0,000151	0,000813	0,000494	0,000782	0,000246	0,000112	1,07E-05	0,000305	0	9,52E-05	0,000178	0,000263	0,002002	0,000377
и	0,000112	0,000378	0,002515	0,000395	0,001439	0,002301	0,000305	0,001775	0,000556	0,001351	0,001693	0,004405	0,002813	0,004179	7,14E-05
й	0	2,38E-06	0	0	0,000177	1,19E-06	0	1,19E-06	0	0	8,21E-05	4,16E-05	6,19E-05	0,000284	1,19E-06
к	0,007803	0	0,00014	0	1,19E-06	0,000481	1,43E-05	3,57E-06	0,002941	0	1,31E-05	0,000325	0	0,000631	0,00786
л	0,007736	2,38E-05	2,38E-06	0,000101	1,31E-05	0,005205	0,000389	1,19E-06	0,00503	0	0,000299	0,000186	0	0,000291	0,005708
м	0,002449	5,95E-06	3,57E-06	2,86E-05	0	0,002934	0	3,57E-06	0,002576	0	5,71E-05	9,4E-05	2,5E-05	0,001203	0,003137
н	0,010834	1,07E-05	5,95E-06	9,4E-05	0,000657	0,009707	1,19E-06	1,43E-05	0,007386	0	0,000234	0	0	0,002812	0,009577
о	0	0,003445	0,006938	0,004562	0,004222	0,002379	0,00188	0,001062	0,000742	0,003024	0,001531	0,005282	0,004817	0,008154	0,000181
п	0,001161	0	0	0	0	0,001982	0	0	0,000727	0	6,07E-05	0,000454	0	7,49E-05	0,008067
р	0,006563	0,000236	0,000305	0,000313	0,000366	0,005307	0,000283	1,78E-05	0,00448	0	0,000539	5,12E-05	0,000132	0,000631	0,006815
с	0,001808	5,59E-05	0,001847	1,43E-05	0,000345	0,004403	2,38E-05	2,26E-05	0,001269	0	0,004964	0,002389	0,000777	0,000977	0,002279
т	0,004958	1,67E-05	0,002716	4,76E-06	0,000103	0,004933	2,38E-06	1,19E-06	0,003844	0	0,000473	0,000247	2,26E-05	0,001085	0,014245
у	3,45E-05	0,000451	0,001096	0,000962	0,001517	0,000237	0,00178	0,000299	2,38E-06	0,000127	0,000719	0,001202	0,001064	0,000195	4,76E-06

➤ для біграм, букви яких не перетинаються

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
а	9,52E-06	0,000692	0,002979	0,000519	0,002198	0,000987	0,001171	0,00405	0,000138	0,00055	0,004176	0,009529	0,002417	0,004456	7,14E-06
б	0,000926	2,38E-06	4,76E-05	7,14E-06	2,86E-05	0,001975	7,14E-06	0	0,000802	0	0,000257	0,000833	7,85E-05	0,000281	0,00187
в	0,005437	1,67E-05	2,38E-05	4,76E-05	0,000259	0,004214	2,38E-06	0,000521	0,004433	0	0,000155	0,000752	0,000138	0,001211	0,006881
г	0,000826	2,38E-06	0	0	0,001218	0,000262	0	0	0,000809	0	0,000152	0,001558	0	0,000159	0,008104
д	0,004295	2,38E-05	0,000804	4,76E-06	4,04E-05	0,004406	1,43E-05	0	0,001994	0	0,00025	0,000804	6,9E-05	0,001549	0,003626
е	4,04E-05	0,001211	0,002298	0,003626	0,002474	0,002118	0,000669	0,001263	7,85E-05	0,002532	0,001692	0,005672	0,003881	0,007005	0,000167
ж	0,001344	6,66E-05	0	1,19E-05	0,000826	0,003876	7,14E-06	0	0,001463	0	9,52E-05	2,38E-06	2,38E-06	0,000975	4,28E-05
з	0,005356	0,00015	0,000857	0,000483	0,000771	0,000231	0,000136	9,52E-06	0,000328	0	9,99E-05	0,000188	0,000271	0,001934	0,000414
и	0,000128	0,000369	0,002505	0,000374	0,001435	0,00231	0,000338	0,001782	0,000564	0,001304	0,001627	0,004402	0,002822	0,004152	6,66E-05
й	0	0	0	0	0,000152	0	0	2,38E-06	0	0	8,57E-05	4,28E-05	5,23E-05	0,000288	2,38E-06
к	0,00779	0	0,000136	0	2,38E-06	0,00045	1,43E-05	0	0,002969	0	1,67E-05	0,000333	0	0,000592	0,007904
л	0,007735	2,38E-05	2,38E-06	8,33E-05	1,43E-05	0,005177	0,000369	2,38E-06	0,005025	0	0,000288	0,000212	0	0,000286	0,005713
м	0,002417	4,76E-06	4,76E-06	3,33E-05	0	0,003012	0	2,38E-06	0,002532	0	7,14E-05	8,8E-05	2,62E-05	0,001275	0,003136
н	0,010769	1,19E-05	7,14E-06	9,99E-05	0,000654	0,009629	2,38E-06	1,9E-05	0,007388	0	0,000221	0	0	0,002779	0,009684
о	0	0,003405	0,006938	0,004668	0,004195	0,002408	0,001849	0,00104	0,000757	0,00311	0,001508	0,005351	0,004804	0,008187	0,000169
п	0,001187	0	0	0	0	0,001977	0	0	0,000719	0	5,47E-05	0,000414	0	7,85E-05	0,008044
р	0,006472	0,000228	0,000295	0,000314	0,000383	0,005434	0,000278	1,19E-05	0,004323	0	0,000562	5,23E-05	0,000136	0,000688	0,006867
с	0,001811	5,47E-05	0,001837	1,67E-05	0,000328	0,004435	1,43E-05	1,67E-05	0,001237	0	0,004877	0,002386	0,000795	0,000983	0,002277
т	0,00483	2,38E-05	0,002693	4,76E-06	9,28E-05	0,004911	2,38E-06	2,38E-06	0,003883	0	0,000454	0,000259	2,38E-05	0,001087	0,014128
у	2,86E-05	0,000428	0,001054	0,000985	0,001435	0,000262	0,001796	0,000305	0	9,99E-05	0,000735	0,001206	0,001016	0,0002	4,76E-06

2. Робота з програмою CoolPinkProgram .

Результати:

$$3,32813 < H(10) < 3,96108$$

Лабораторная работа №1

×

Произвольная часть текста:
ми_но_тот

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:

Символ по счету:

Номер эксперимента: 51

Поле ввода символов:

Продолжить

Другой

Неравенство для энтропии:
3,32812539321919< H < 3,961076880026

Двоичная таблица угаданных символов:
01000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
00000000010000000000000000000000
01000000000000000000000000000000
00000000000000000001000000000000

Вероятности:

q[1] = 0.24
q[2] = 0.1
q[3] = 0.02
q[4] = 0.08
q[5] = 0.04
q[6] = 0.04
q[7] = 0.02
q[8] = 0.02
q[9] = 0.04
q[10] = 0.06
q[11] = 0.02
q[12] = 0.04
q[13] = 0.04
q[14] = 0.02
q[15] = 0.06
q[16] = 0
q[17] = 0.02
q[18] = 0
q[19] = 0.02
q[20] = 0
q[21] = 0.02
q[22] = 0
q[23] = 0
q[24] = 0
q[25] = 0
q[26] = 0.02
q[27] = 0.04
q[28] = 0
q[29] = 0.02
q[30] = 0.02
q[31] = 0
q[32] = 0

$$2,1815722 < H(20) < 2,80936$$

Лабораторная работа №1

×

Произвольная часть текста:
дуг_попасться_инди

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:

Символ по счету:

Номер эксперимента: 52

Поле ввода символов:

Продолжить

Другой

Неравенство для энтропии:
2,18157222774133< H < 2,80936468627839

Двоичная таблица угаданных символов:
00000000010000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
01000000000000000000000000000000
00000000000000000000000000000000

Вероятности:

q[1] = 0.4509803
q[2] = 0.1960784
q[3] = 0
q[4] = 0.0196078
q[5] = 0.0196078
q[6] = 0.0196078
q[7] = 0
q[8] = 0
q[9] = 0.0196078
q[10] = 0.019607
q[11] = 0
q[12] = 0.039215
q[13] = 0.019607
q[14] = 0.058823
q[15] = 0
q[16] = 0.019607
q[17] = 0
q[18] = 0.039215
q[19] = 0
q[20] = 0
q[21] = 0.019607
q[22] = 0
q[23] = 0
q[24] = 0.019607
q[25] = 0
q[26] = 0.019607
q[27] = 0
q[28] = 0.019607
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0

Строка состояния:

$$1,77114 < H(30) < 2,4520$$

Лабораторная работа №1

Произвольная часть текста:
вам_то_мое_место_я_его_первый

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:

Символ по счету:

Номер эксперимента: 50

Поле ввода символов:

Продолжить Другой

Неравенство для энтропии:
1,77113965312456 < H < 2,45204727688637

Двоичная таблица угаданных символов:

00000010000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:

q[1] = 0,5714285
q[2] = 0,1020408
q[3] = 0,0612244
q[4] = 0
q[5] = 0,0408163
q[6] = 0
q[7] = 0,0204081
q[8] = 0,0204081
q[9] = 0,0204081
q[10] = 0
q[11] = 0
q[12] = 0
q[13] = 0
q[14] = 0
q[15] = 0
q[16] = 0
q[17] = 0,020408
q[18] = 0
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0,040816
q[24] = 0,020408
q[25] = 0
q[26] = 0,020408
q[27] = 0,020408
q[28] = 0,020408
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0,020408

Строка состояния:

3. Використані формули у роботі:

- Для обчислення надлишковості:

$$R = 1 - \frac{H_{\infty}}{H_0}$$

$$H_0 = \log_2 m,$$

m - к-ть букв у алфавіті

- Для обчислення ентропій H1 і H2

$$H_n = \frac{1}{n} H(x_1, x_2, \dots, x_n)$$

, де

$$H(x_1, x_2, \dots, x_n) = - \sum_{z_1, z_2, \dots, z_n} P(x_1 = z_1, \dots, x_n = z_n) \cdot \log_2 P(x_1 = z_1, \dots, x_n = z_n).$$

Отримані результати ентропій та надлишковостей занесли у таблицю:

	Ентропія	Надлишковість
Букви, текст з пробілами	4.352138	0.144536
Букви, текст без пробілів	4.444298	0.11896
Біграми, що не перетин., текст з пробілами	3.942818	0.22499
Біграми, що не перетин., текст без пробілів	4.125403	0.18218
Біграми, що перетин., текст з пробілами	3.94275	0.225006
Біграми, що перетин., текст без пробілів	4.12421	0.182416

$3,32813 < H(10) < 3,96108$	$0,2147 < R(10) < 0,340258$
$2,1815722 < H(20) < 2,80936$	$0,443144 < R(20) < 0,567639$
$1,77114 < H(30) < 2,4520$	$0,513916 < R(30) < 0,648917$

2. Аналіз результатів

Ми ознайомилися з такими поняттями і ось як ми їх зрозуміли:

- ентропія - це те, скільки інформації про джерело нам невідомо.
- надлишковість - величина, що показує, на скільки коротшим може бути повідомлення, при чому зміст залишається незмінним.

Проаналізувавши отримані значення частот, ми переконалися в тому, що найчастішими літерами в російському алфавіті є:

Table of frequencies:

о	<0.11472>
е	<0.08683>
а	<0.08321>
н	<0.06909>
и	<0.06640>
т	<0.06030>
с	<0.05361>
л	<0.05023>
в	<0.04649>
р	<0.03950>

Отримані значення ентропій, показують те, що вона більша у тексті без пробілів.

Висновки

У результаті виконання практикума ми ознайомилися з поняттями ентропії на символ джерела та його надлишковості, набули практичних навичок щодо оцінки ентропії та надлишковості.