НТУУ "КПІ ім Ігоря Сікорського" Фізико-технічний інститут

КРИПТОГРАФІЯ КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконали:

студенти групи ФБ-14 Разумний Ілля Болгов Микола

Перевірила:

Селюх П.В.

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

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 1Mб), де імовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 на тому ж тексті, в якому вилучено всі пробіли

Обраний текст міститься у файлі <u>ds.txt</u> Програма <u>razik_bolgov.py</u> на мові Python3

Частота букв в тексті з пробілом:

```
Letter frequencies in the text with spaces
      0.15158885306697925
    0.07048301060088764
"e" | 0.06929070790229701
 "н" | 0.051603665416455606
     0.050129745209302155
         0.04580124752898923
     0.04331606445017839
"B" | 0.04018535552381771
"p" | 0.03990556670037543
    0.02629832071845381
         0.026206886462426922
    0.025910639472899802
    0.024559241168822382
"ы" | 0.016580687987916048
     0.015013504839615171
 "ь"
"3" | 0.013969325635788099
 "ж"
         0.007815800205178471
     0.0045808562269471385
       0.004368728752964755
         0.0020883584076541446
         0.001709820587702824
```

Ентропія з пробілом:

H.1 = 4.403827131527339

Надлишковість з пробілом:

R.1 = 0.11923457369453216

Частота букв в тексті без пробіла (ми не враховуємо пробіл як частину алфавіта):

```
Letter frequencies in the text without spaces
     0.10734870006423161
"a" | 0.08307647872812784
"e" | 0.08167114276230424
"и" | 0.07104921693472947
"H" | 0.060823888987080396
    | 0.059086617839145074
"c" | 0.05398473098162286
    0.051055510770650035
"в"
         0.04736542614873283
         0.04703564638988158
         0.03591581778913926
         0.03099714190875662
         0.03088937074573334
         0.030540192177537902
         0.02894733438805378
         0.019543222702642117
         0.01886210895233497
         0.017696024968423048
         0.017092506455492665
         0.01646527828669716
         0.014978036236975856
         0.012462657292012432
         0.010186530328960698
         0.00921227901523022
         0.008494523069495157
         0.0053993352674664725
         0.005149306169252456
         0.0033904807887124794
         0.002461493363451781
      0.00201532074853539
```

Ентропія без пробіла:

H.1 = 4.467203167107593

Надлишковість без пробіла:

R.1 = 0.0982991211426687

Частота біграм, що перетинаються, в тексті з пробілами: bigram_table_cross_spaces.xlsx

	a	6	В	r	д	e	ж	3	И	й	K	л	м	н	0	п	р	c	т	У	Φ	x	ц	ч	ш	щ	ы	ь	9	ю	я	
а	7.31E-06	0.000739	0.003059	0.00066	0.002183	0.000997	0.001287	0.002864	0.000133	0.000892	0.003502	0.008564	0.003365	0.004113	1.1E-05	0.001918	0.003526	0.00376	0.005528	0.000165	0.000263	0.001317	0.00032	0.000913	0.000966	0.000227	0	0	1.28E-05	0.000719	0.001986	0.016487
6	0.000975	1.1E-05	4.39E-05	3.66E-06	1.1E-05	0.002461	3.11E-05	7.31E-06	0.000936	0	9.33E-05	0.001037	3.84E-05	0.000258	0.002092	1.28E-05	0.001317	0.00015	1.28E-05	0.001009	0	3.11E-05	1.65E-05	7.31E-06	3.47E-05	0.000163	0.002791	0.000432	0	7.86E-05	0.000223	0.000225
В	0.006258	1.28E-05	2.93E-05	2.38E-05	0.000415	0.005665	1.83E-06	0.000336	0.003703	0	0.000287	0.000808	9.87E-05	0.000956	0.006411	0.000161	0.000616	0.001966	0.000358	0.000739	0	0.000106	6.95E-05	7.13E-05	0.000801	1.28E-05	0.003127	8.96E-05	0	0	0.000274	0.006786
r	0.001489	1.83E-06	1.46E-05	0	0.000724	0.000316	0	1.83E-06	0.000695	0	0.000139	0.001035	7.31E-06	0.000219	0.006847	3.66E-06	0.001487	3.47E-05	7.31E-06	0.000658	0	0	0	9.14E-06	1.1E-05	0	0	0	3.66E-06	0	0	0.000552
А	0.004285	3.84E-05	0.001315	1.1E-05	4.39E-05	0.005044	9.69E-05	1.46E-05	0.002361	0	0.000318	0.000525	0.000123	0.001642	0.003888	0.000152	0.000927	0.000441	8.59E-05	0.001434	0	4.57E-05	0.000424	3.11E-05	7.13E-05	0	0.000576	0.000426	0	3.47E-05	0.000408	0.00115
e	0.000163	0.001019	0.002734	0.002379	0.002964	0.001754	0.000883	0.001335	0.000144	0.002213	0.001372	0.006183	0.003206	0.006969	0.000219	0.00079	0.006239	0.004771	0.004736	0.000207	4.39E-05	0.000594	0.000556	0.000977	0.000664	0.000649	0	0	0	0.000106	0.000197	0.015224
ж	0.001432	3.29E-05	0	1.46E-05	0.000772	0.003165	7.31E-06	0	0.001483	0	0.000194	2.38E-05	7.31E-06	0.000931	6.4E-05	0	3.66E-06	1.28E-05	3.66E-06	0.00024	0	0	1.83E-06	4.75E-05	0	0	0	2.56E-05	0	0	0	0.000181
3	0.005298	0.000106	0.000817	0.000183	0.000794	0.000561	6.77E-05	1.1E-05	0.000499	0	9.51E-05	0.000216	0.000318	0.001421	0.000667	3.66E-06	0.000192	3.84E-05	1.28E-05	0.000346	0	0	7.31E-06	6.58E-05	7.31E-06	0	0.000353	0.000124	0	1.83E-06	0.000271	0.001492
И	0.00023	0.000342	0.00229	0.000614	0.001573	0.001922	0.000274	0.002213	0.000722	0.001452	0.002862	0.004945	0.002255	0.003981	0.000552	0.001165	0.001159	0.003259	0.004758	5.49E-05	9.33E-05	0.001478	0.000991	0.002631	0.000466	0.00034	0	0	1.83E-06	0.000291	0.00103	0.016334
й	0	5.49E-06	0	1.83E-06	0.000159	5.49E-06	0	1.65E-05	0	0	0.00017	0.000124	5.49E-05	0.000375	1.28E-05	0	1.83E-06	0.000521	0.000421	0	1.83E-06	1.83E-06	5.12E-05	0.000199	9.14E-05	5.49E-06	0	0	0	0	0	0.008353
K	0.006506	0	0.000366	3.66E-06	9.14E-06	0.000726	1.1E-05	8.23E-05	0.003826	0	2.74E-05	0.000856	2.01E-05	0.000395	0.008615	7.31E-06	0.00207	0.000243	0.000693	0.001818	0	1.83E-06	0.00019	1.83E-06	1.1E-05	0	0	0	0	3.66E-06	0	0.003988
л	0.004883	6.58E-05	2.74E-05	0.000155	4.94E-05	0.004661	0.00024	2.38E-05	0.007783	0	0.000492	0.000287	7.31E-06	0.00039	0.005806	6.95E-05	1.83E-06	0.001686	0.000143	0.001403	7.31E-06	0	0	0.000148	3.66E-06	3.66E-06	0.000653	0.00425	0	0.000832	0.001609	0.007637
M	0.003535	0.000294	0.000117	1.65E-05	0	0.003367	0	1.83E-06	0.003195	0	0.000113	0.000174	9.87E-05	0.001103	0.003354	0.000177	0.000101	0.000148	1.83E-05	0.001841	4.02E-05	1.46E-05	3.11E-05	4.21E-05	1.1E-05	3.66E-06	0.000852	0.00011	0	0	0.000402	0.007137
н	0.008984	1.83E-05	1.65E-05	0.000159	0.000708	0.007328	0.000143	3.29E-05	0.006786	0	0.000479	3.66E-06	0	0.002679	0.009445	1.83E-06	6.77E-05	0.000742	0.000922	0.00273	2.01E-05	5.49E-06	0.000596	0.000267	9.14E-06	5.85E-05	0.004135	0.001081	9.14E-06	0.000104	0.001086	0.002986
0	1.83E-05	0.003348	0.008	0.00399	0.005201	0.001543	0.001876	0.001196	0.000722	0.003767	0.00203	0.006673	0.005499	0.00508	0.000179	0.00141	0.006821	0.0067	0.005385	9.33E-05	0.000133	0.000625	0.000154	0.001831	0.000898	0.000267	0	0	7.86E-05	0.000216	0.000569	0.016771
п	0.002017	0	0	0	1.83E-06	0.002425	0	0	0.00141	0	6.4E-05	0.000836	5.49E-06	0.000137	0.009655	0.001004	0.005824	7.13E-05	0.000124	0.000849	5.49E-06	0	7.31E-06	1.46E-05	1.1E-05	5.49E-06	0.000304	6.4E-05	0	3.66E-06	0.000432	0.000936
р	0.007838	9.87E-05	0.000369	0.00036	0.000315	0.005696	0.000347	0.000115	0.005062	0	0.000413	0.000108	0.00036	0.00134	0.007847	0.000115	5.67E-05	0.000582	0.000706	0.002675	5.12E-05	0.000179	7.86E-05	8.78E-05	0.000179	5.49E-05	0.001518	0.000508	0	0.000174	0.000894	0.001776
C	0.001812	7.68E-05	0.001304	4.02E-05	0.000176	0.002933	9.14E-06	7.31E-06	0.002017	0	0.004345	0.002306	0.000795	0.000958	0.002644	0.001772	0.000254	0.00096	0.012014	0.000653	1.65E-05	0.000174	6.03E-05	0.000294	7.31E-05	0	0.000419	0.002913	0	0.000252	0.00367	0.002851
T	0.0067	1.65E-05	0.002776	5.49E-06	0.000214	0.005883	0	1.83E-06	0.00382	0	0.00056	0.000216	2.01E-05	0.001282	0.009646	7.68E-05	0.003357	0.000962	6.4E-05	0.002244	5.49E-06	2.38E-05	0.000113	9.69E-05	9.14E-06	3.66E-05	0.001565	0.004467	0	0.000126	0.000576	0.005265
y	8.59E-05	0.001064	0.000503	0.000744	0.001659	0.000249	0.001264	0.0004	1.83E-05	0.000139	0.001022	0.002264	0.000942	0.000399	2.38E-05	0.000728	0.00092	0.00126	0.001346	5.49E-06	3.84E-05	0.000528	1.83E-05	0.000812	0.00068	0.000225	0	0	1.46E-05	0.001024	8.23E-05	0.0061
ф	0.000236	0	0	0	0	0.000457	0	1.83E-06	0.000283	0	1.83E-06	7.31E-05	7.31E-06	1.1E-05	0.000309	0	8.78E-05	2.01E-05	2.19E-05	8.41E-05	1.65E-05	0	0	0	0	0	1.1E-05	2.56E-05	0	3.66E-06	0	5.85E-05
x	0.000735	0	0.000181	3.66E-06	0	0.000113	0	0	0.000294	0	9.14E-06	0.000183	0.000135	0.00023	0.002193	0	0.000132	6.4E-05	3.29E-05	0.000172	0	5.49E-06	1.83E-06	0	1.28E-05	0	0	1.83E-06	7.31E-06	0	0	0.003308
ц	0.000949	0	0.00015	0	0	0.001024	0	0	0.000763	0	0.000205	0	5.49E-06	0	0.000192	0	0	5.49E-06	0	0.000274	1.83E-06	1.83E-06	1.83E-06	0	0	0	0.000276	0	0	0	0	0.000519
ч	0.002209	0	7.31E-06	0	0	0.003308	1.83E-06	0	0.001671	0	0.000675	6.22E-05	3.66E-06	0.000858	0.000101	0	5.3E-05	0	0.00175	0.000574	0	0	0	0	7.13E-05	0	0	0.000216	0	0	0	0.001147
ш	0.001022	0	3.11E-05	0	0	0.001993	0	0	0.001622	0	0.000461	0.000518	1.28E-05	0.000386	0.000305	2.56E-05	1.83E-06	1.83E-06	0.000139	0.000331	0	1.1E-05	0	0	0	0	0	0.000247	0	5.49E-06	0	9.33E-05
щ	0.000464	0	1.83E-06	0	0	0.001225	0	0	0.000828	0	0	0	0	5.67E-05	0	0	0	0	0	0.000143	0	0	0	0	0	0	0	3.29E-05	0	0	0	0.000124
ы	0	0.000161	0.000865	0.000144	0.000141	0.001379	3.29E-05	6.22E-05	4.21E-05	0.00198	0.000223	0.001858	0.001333	0.000207	0	0.000205	0.00026	0.000675	0.000618	1.83E-06	0	0.001152	5.49E-06	0.000132	0.000448	1.83E-05	0	0	0	0	1.46E-05	0.004623
b	0	8.41E-05	2.01E-05	0.00019	0.000157	0.000658	0	0.000104	7.68E-05	0	0.000834	0	0.000196	0.0013	6.03E-05	5.49E-06	3.66E-06	0.000688	0.00019	0	3.29E-05	5.3E-05	9.69E-05	5.3E-05	0.000545	1.83E-05	0	0	0	0.000395	0.000913	0.008339
9	0	0	0	0	3.66E-06	0	0	1.83E-06	0	1.83E-06	8.78E-05	0.000144	8.05E-05	2.56E-05	1.83E-06	3.47E-05	6.58E-05	4.75E-05	0.001551	0	2.01E-05	2.01E-05	0	0	1.83E-06	0	0	0	0	0	0	0
ю	3.66E-06	0.000304	5.49E-06	1.65E-05	0.000315	0	2.93E-05	4.75E-05	0	5.49E-06	0.000165	2.38E-05	8.59E-05	6.58E-05	0	7.31E-06	7.13E-05	0.000104	0.000399	0	0	1.28E-05	3.11E-05	0.000123	2.56E-05	0.000413	0	0	0	3.29E-05	1.83E-06	0.002293
Я	0	0.000108	0.000247	9.51E-05	0.00041	9.51E-05	0.000172	0.00019	3.29E-05	0.000108	0.000201	0.000885	0.0004	0.000955	0	0.000106	0.000197	0.000594	0.001417	1.83E-06	5.49E-06	0.000177	7.13E-05	0.000203	2.19E-05	0.000256	0	0	0	0.000113	8.78E-05	0.008851
	0.002348	0.006554	0.014895	0.00444	0.006914	0.002335	0.001867	0.004901	0.009346	1.46E-05	0.009032	0.002386	0.006817	0.012881	0.009933	0.016255	0.004091	0.015293	0.006671	0.003815	0.000913	0.001256	0.000474	0.00365	0.001083	0.000119	1.83E-06	0	0.00196	6.4E-05	0.001276	0

Ентропія для біграм, що перетинаються, з пробілами:

H.2 = 4.021115482275827

Надлишковість для біграм, що перетинаються, з пробілами:

R.2 = 0.19577690354483457

Частота біграм, що не перетинаються, в тексті з пробілами:

bigram_table_not_cross_spaces.xlsx

	a	6	В	r	Д	e	ж	3	И	й	K	Л	M	н	0	n	р	c	T	у	φ	x	ц	ч	ш	щ	ы	ь	9	ю	я	
a	7.31E-06	0.000735	0.003138	0.000636	0.002132	0.001031	0.001291	0.002867	0.000128	0.000889	0.003581	0.00865	0.003299	0.003994	3.66E-06	0.001895	0.003431	0.003906	0.005523	0.000197	0.000285	0.00132	0.000304	0.000947	0.000984	0.000245	0	0	2.19E-05	0.000812	0.001935	0.016235
6	0.00102	1.46E-05	5.12E-05	3.66E-06	3.66E-06	0.002377	3.29E-05	1.1E-05	0.000896	0	9.14E-05	0.000969	4.39E-05	0.000267	0.002074	1.83E-05	0.00128	0.000176	1.1E-05	0.000977	0	2.56E-05	1.1E-05	1.46E-05	4.02E-05	0.000168	0.002794	0.000464	0	6.95E-05	0.000223	0.000241
В	0.006071	1.46E-05	2.93E-05	2.93E-05	0.000402	0.00576	0	0.000296	0.003687	0	0.000293	0.000786	0.000102	0.000962	0.006485	0.000154	0.000607	0.001949	0.000362	0.000721	0	9.51E-05	5.85E-05	8.41E-05	0.000779	1.1E-05	0.003072	8.05E-05	0	0	0.000252	0.006759
г	0.001489	0	7.31E-06	0	0.000735	0.000307	0	3.66E-06	0.000731	0	0.000132	0.001031	7.31E-06	0.000256	0.006726	3.66E-06	0.001529	3.66E-05	0	0.000622	0	0	0	1.1E-05	7.31E-06	0	0	0	0	0	0	0.000556
А	0.004177	3.29E-05	0.001353	1.83E-05	3.66E-05	0.005102	8.41E-05	1.83E-05	0.002381	0	0.0003	0.000508	0.000121	0.001671	0.003943	0.000172	0.000856	0.000395	8.78E-05	0.001481	0	3.29E-05	0.000417	2.56E-05	6.58E-05	0	0.000574	0.00045	0	4.75E-05	0.000413	0.00117
e	0.000194	0.001006	0.002611	0.002418	0.003058	0.001829	0.000929	0.001353	0.000124	0.002198	0.001382	0.006305	0.0032	0.007084	0.000227	0.000786	0.006228	0.004736	0.004674	0.000238	4.02E-05	0.000647	0.0006	0.000984	0.000582	0.000658	0	0	0	0.000102	0.000216	0.015076
ж	0.001415	3.66E-05	0	1.83E-05	0.000786	0.003101	1.1E-05	0	0.001562	0	0.00019	1.83E-05	7.31E-06	0.00094	5.49E-05	0	0	1.46E-05	3.66E-06	0.000238	0	0	0	3.66E-05	0	0	0	1.83E-05	0	0	0	0.000176
3	0.005468	0.000102	0.00083	0.00019	0.000812	0.000571	7.68E-05	1.83E-05	0.000443	0	0.000113	0.000216	0.00034	0.001393	0.000647	0	0.000179	2.93E-05	1.83E-05	0.000373	0	0	3.66E-06	6.95E-05	1.46E-05	0	0.000358	0.000139	0	0	0.000278	0.00158
И	0.000219	0.000329	0.002187	0.000614	0.001584	0.002033	0.000278	0.00211	0.000768	0.001514	0.00271	0.004981	0.002264	0.004034	0.000527	0.00117	0.001189	0.003233	0.004634	5.12E-05	6.58E-05	0.001478	0.001024	0.002546	0.000472	0.000391	0	0	0	0.000318	0.001057	0.016308
й	0	7.31E-06	0	3.66E-06	0.000135	1.1E-05	0	7.31E-06	0	0	0.000165	0.000143	5.85E-05	0.000344	1.46E-05	0	0	0.00056	0.000421	0	0	0	4.39E-05	0.000227	7.68E-05	3.66E-06	0	0	0	0	0	0.008138
K	0.006737	0	0.000355	3.66E-06	3.66E-06	0.00071	3.66E-06	7.68E-05	0.003756	0	3.29E-05	0.000896	2.19E-05	0.000384	0.008474	7.31E-06	0.002048	0.000256	0.000655	0.001803	0	0	0.000197	3.66E-06	1.46E-05	0	0	0	0	3.66E-06	0	0.004089
Л	0.004839	6.58E-05	3.29E-05	0.000143	4.75E-05	0.004546	0.000249	2.19E-05	0.007867	0	0.000497	0.000289	7.31E-06	0.000391	0.005647	7.31E-05	0	0.001635	0.000132	0.001339	7.31E-06	0	0	0.00015	3.66E-06	3.66E-06	0.000658	0.004279	0	0.000768	0.001591	0.007688
M	0.003654	0.000285	9.87E-05	1.1E-05	0	0.003325	0	0	0.003167	0	0.000106	0.000172	7.31E-05	0.001068	0.003394	0.000201	7.31E-05	0.000157	2.19E-05	0.001938	4.39E-05	7.31E-06	3.29E-05	4.02E-05	1.46E-05	3.66E-06	0.000874	0.000132	0	0	0.000402	0.007198
н	0.009078	1.1E-05	1.46E-05	0.000168	0.000695	0.007234	0.000132	4.75E-05	0.006536	0	0.000446	7.31E-06	0	0.00271	0.009586	3.66E-06	6.95E-05	0.000768	0.00094	0.002688	1.83E-05	0	0.000618	0.000194	7.31E-06	5.12E-05	0.004056	0.001123	1.1E-05	0.000124	0.001075	0.003039
0	1.46E-05	0.003368	0.0082	0.004019	0.005219	0.001503	0.001847	0.00113	0.000706	0.003928	0.002037	0.006821	0.005289	0.005098	0.000187	0.001339	0.006847	0.006715	0.005417	8.05E-05	0.000143	0.0006	0.00015	0.00196	0.000907	0.000271	0	0	8.05E-05	0.000187	0.000519	0.016718
п	0.001979	0	0	0	0	0.002436	0	0	0.001467	0	5.85E-05	0.000841	7.31E-06	0.000132	0.009557	0.000984	0.005819	6.22E-05	0.000139	0.000885	1.1E-05	0	3.66E-06	1.83E-05	1.1E-05	1.1E-05	0.000296	5.49E-05	0	0	0.000413	0.000859
р	0.007684	9.87E-05	0.000413	0.000369	0.000296	0.005698	0.000344	8.05E-05	0.005179	0	0.000424	0.000128	0.000369	0.001372	0.007801	0.00011	5.49E-05	0.000592	0.000761	0.002699	6.58E-05	0.000183	6.95E-05	8.41E-05	0.000194	5.12E-05	0.001591	0.000494	0	0.000157	0.00087	0.001803
С	0.001876	8.41E-05	0.001254	4.39E-05	0.000165	0.002897	7.31E-06	3.66E-06	0.001971	0	0.004513	0.002322	0.000764	0.00094	0.002688	0.001719	0.000274	0.000885	0.012007	0.000655	1.46E-05	0.000168	6.58E-05	0.000315	6.22E-05	0	0.000402	0.002886	0	0.000234	0.003624	0.002805
T	0.0067	1.46E-05	0.002725	1.1E-05	0.00023	0.005691	0	3.66E-06	0.003822	0	0.000618	0.000245	1.83E-05	0.001258	0.009597	9.14E-05	0.003412	0.001028	7.31E-05	0.002151	3.66E-06	2.93E-05	9.14E-05	8.05E-05	1.1E-05	4.39E-05	0.001445	0.004579	0	0.000106	0.000596	0.005332
У	7.31E-05	0.001141	0.000501	0.000699	0.001635	0.000263	0.001254	0.00041	1.83E-05	0.000132	0.001028	0.00229	0.001009	0.000399	2.93E-05	0.000786	0.000944	0.00124	0.001441	3.66E-06	3.66E-05	0.000519	2.19E-05	0.000772	0.000739	0.000223	0	0	7.31E-06	0.001031	6.95E-05	0.005907
ф	0.000208	0	0	0	0	0.000428	0	3.66E-06	0.000252	0	0	8.78E-05	3.66E-06	1.46E-05	0.000347	0	7.68E-05	2.19E-05	2.56E-05	5.85E-05	1.46E-05	0	0	0	0	0	1.1E-05	3.29E-05	0	3.66E-06	0	5.12E-05
x	0.000728	0	0.000168	7.31E-06	0	0.000124	0	0	0.000315	0	1.46E-05	0.000183	0.000132	0.000252	0.002227	0	0.000132	5.49E-05	2.93E-05	0.000172	0	3.66E-06	3.66E-06	0	1.1E-05	0	0	0	7.31E-06	0	0	0.003248
ц	0.001031	0	0.000179	0	0	0.001006	0	0	0.00075	0	0.000168	0	3.66E-06	0	0.00019	0	0	1.1E-05	0	0.000296	3.66E-06	0	0	0	0	0	0.000271	0	0	0	0	0.000483
ч	0.002132	0	0	0	0	0.003325	3.66E-06	0	0.001734	0	0.000644	4.02E-05	3.66E-06	0.000841	0.000106	0	5.85E-05	0	0.001726	0.000545	0	0	0	0	7.31E-05	0	0	0.000194	0	0	0	0.00117
W	0.001064	0	3.66E-05	0	0	0.00203	0	0	0.001657	0	0.000417	0.000574	1.1E-05	0.000377	0.0003	2.19E-05	0	3.66E-06	0.000121		0	1.1E-05	0	0	0	0	0	0.000274	0	3.66E-06	0	8.05E-05
щ	0.00041	0	0	0		0.001156	0		0.000881	0	0	0	-	5.49E-05	0	0	0	0		0.000124	0	0	0	0	0	0	0	3.29E-05	0	0	0	0.000102
ы	0	0.000172	0.000922	0.000124	0.000143	0.001423	3.66E-05	5.12E-05	3.66E-05	0.00199	0.000241					0.000187	0.000263	0.000706	0.000629	0	0	0.001141	7.31E-06	0.000117	0.000424	2.19E-05	0	0	0	0	1.46E-05	0.004729
b	0	7.31E-05	2.93E-05	0.000179	0.000161	0.000699	0	0.000102	7.68E-05	0	0.000852	0	0.000208	0.001276	6.22E-05	3.66E-06	0	0.000644	0.000176	0	5.12E-05	5.12E-05	8.78E-05	4.02E-05	0.000519	1.83E-05	0	0	0	0.00041	0.000856	0.008218
9	0	0	0	0	3.66E-06	0	0	0	0	3.66E-06	8.78E-05	0.000165	9.14E-05	2.56E-05	0	3.29E-05	6.95E-05	4.75E-05	0.001562	0	2.19E-05	1.83E-05	0	0	3.66E-06	0	0	0	0	0	0	0
ю	3.66E-06	0.000336	1.1E-05	1.83E-05	0.0003	0	2.56E-05	5.85E-05	0	3.66E-06	0.000179	2.93E-05	0.00011	6.58E-05	0	3.66E-06	9.14E-05	0.00011	0.000333	0	0	1.1E-05	4.39E-05	0.000146	2.56E-05	0.000439	0	0	0	3.29E-05	3.66E-06	0.002198
я	0	0.000106	0.000252	8.05E-05	0.000424	0.000106	0.000139	0.000219	3.66E-05	0.000117	0.000205	0.000856	0.000424	0.001017	0	9.14E-05	0.000194	0.000582	0.001419	3.66E-06	0	0.000172	8.05E-05	0.000267	1.83E-05	0.000274	0	0	0	0.000102	9.87E-05	0.008876
	0.002271	0.006598	0.015068	0.004513	0.00688	0.002374	0.001913	0.00478	0.00952	1.46E-05	0.008884	0.002311	0.006766	0.0129	0.009959	0.016517	0.004049	0.015401	0.006912	0.003859	0.000951	0.001306	0.00041	0.003687	0.001072	0.000102	3.66E-06	0	0.001916	6.95E-05	0.001335	0

Ентропія біграм, що не перетинаються, в тексті з пробілами:

H.2 = 4.020669087731644

Надлишковість для біграм, що не перетинаються, з пробілами:

R.2 = 0.19586618245367116

Частота біграм, що перетинаються, в тексті без пробілів: bigram_table_cross_not_spaces.xlsx

	a	6	В	r	Д	e	ж	3	и	й	К	л	M	н	0	n	р	С	T	У	Φ	x	ц	ч	ш	щ	ы	ь	9	ю	я
a	0.000371	0.001636	0.005557	0.001323	0.003406	0.001502	0.001711	0.003975	0.001414	0.001052	0.00521	0.010374	0.005009	0.006473	0.001136	0.004328	0.004636	0.006568	0.007408	0.000679	0.000481	0.001681	0.000418	0.001461	0.001257	0.000287	0	0	0.000328	0.000856	0.002541
6	0.001155	1.29E-05	6.9E-05	8.62E-06	1.94E-05	0.00291	3.66E-05	1.72E-05	0.001142	0	0.000119	0.001226	4.96E-05	0.000325	0.002498	2.8E-05	0.001556	0.00019	2.37E-05	0.001201	4.31E-06	3.66E-05	1.94E-05	1.94E-05	4.31E-05	0.000192	0.003289	0.000509	3.02E-05	9.27E-05	0.000269
В	0.007458	0.000293	0.000463	0.000399	0.000875	0.006746	8.62E-05	0.000608	0.00466	2.16E-06	0.001121	0.001101	0.000405	0.001711	0.008022	0.001048	0.001043	0.003319	0.000976	0.00106	3.88E-05	0.000175	0.000144	0.000239	0.001013	2.37E-05	0.003686	0.000106	0.000175	4.31E-06	0.000362
r	0.001772	2.8E-05	7.11E-05	1.29E-05	0.000892	0.000382	6.47E-06	3.02E-05	0.000877	0	0.000194	0.001226	1.72E-05	0.000366	0.008113	7.11E-05	0.001761	9.48E-05	2.16E-05	0.000804	0	2.16E-06	0	3.23E-05	1.51E-05	0	0	0	8.62E-06	0	4.31E-06
А	0.005076	0.000103	0.001688	4.53E-05	0.000108	0.005953	0.000121	6.47E-05	0.002865	0	0.000491	0.000655	0.000177	0.002084	0.004664	0.000293	0.001138	0.000655	0.000181	0.001735	2.16E-06	6.9E-05	0.0005	5.17E-05	9.7E-05	0	0.000679	0.000502	1.51E-05	4.1E-05	0.000483
e	0.000394	0.002201	0.004865	0.003317	0.004397	0.002326	0.001272	0.002218	0.001088	0.002608	0.002552	0.00768	0.004559	0.009689	0.001319	0.00294	0.00777	0.007462	0.006415	0.000819	0.000116	0.000877	0.000692	0.001608	0.000953	0.000782	0	0	0.000241	0.000125	0.000386
ж	0.001688	4.53E-05	2.16E-05	2.16E-05	0.000916	0.003735	8.62E-06	1.29E-05	0.001755	0	0.000241	3.02E-05	1.51E-05	0.001127	8.62E-05	2.8E-05	1.72E-05	3.23E-05	1.29E-05	0.000285	6.47E-06	2.16E-06	2.16E-06	5.82E-05	0	0	0	3.02E-05	2.16E-06	0	6.47E-06
3	0.006266	0.000196	0.001121	0.000304	0.001073	0.000677	9.91E-05	0.000112	0.000629	0	0.000259	0.000272	0.000466	0.00177	0.000892	0.000244	0.000261	0.0002	9.91E-05	0.000442	1.08E-05	2.16E-06	2.16E-05	0.000123	1.72E-05	0	0.000416	0.000147	1.51E-05	2.16E-06	0.00033
И	0.000502	0.001276	0.004684	0.001216	0.002701	0.002578	0.000463	0.003289	0.002037	0.001714	0.004414	0.006158	0.003268	0.006466	0.002011	0.003604	0.001936	0.005891	0.006389	0.000558	0.000213	0.001884	0.001248	0.003524	0.000685	0.000429	2.16E-06	0	0.000198	0.000351	0.00136
й	0.00019	0.00041	0.000696	0.000422	0.000767	0.000101	0.000188	0.000336	0.000632	0	0.000981	0.000336	0.000498	0.001039	0.000528	0.001048	0.000304	0.001638	0.000907	0.000203	7.76E-05	0.000101	0.000119	0.00053	0.000209	1.29E-05	0	0	9.05E-05	1.51E-05	8.41E-05
К	0.007753	0.000237	0.000929	0.000134	0.000246	0.000914	0.000112	0.000237	0.004813	0	0.000269	0.001067	0.000183	0.00086	0.010473	0.000444	0.002539	0.000817	0.001071	0.002231	6.47E-06	3.66E-05	0.000241	0.000127	3.02E-05	0	0	0	8.19E-05	6.47E-06	5.6E-05
Л	0.005912	0.000375	0.001088	0.000446	0.000358	0.005708	0.000319	0.000261	0.009949	2.16E-06	0.001073	0.000429	0.000209	0.001198	0.008191	0.000877	0.0002	0.002755	0.000466	0.001806	1.72E-05	5.6E-05	1.72E-05	0.000461	6.04E-05	1.08E-05	0.000769	0.005009	8.62E-05	0.000981	0.001964
M	0.004337	0.000677	0.000942	0.000364	0.000399	0.004076	0.000123	0.000259	0.004358	0	0.000629	0.000325	0.000511	0.00197	0.00449	0.001147	0.000345	0.000987	0.000321	0.002349	8.62E-05	9.27E-05	5.39E-05	0.000248	0.000103	6.47E-06	0.001004	0.000129	9.27E-05	4.31E-06	0.000567
н	0.010622	0.000198	0.000403	0.000256	0.000994	0.008695	0.000207	0.000164	0.008195	0	0.000737	3.88E-05	0.000129	0.003464	0.011353	0.000466	0.000172	0.001287	0.001209	0.003352	2.59E-05	4.31E-05	0.000707	0.000401	4.31E-05	7.11E-05	0.004873	0.001274	3.02E-05	0.000123	0.001291
0	0.000259	0.004975	0.011508	0.005229	0.007042	0.002162	0.002548	0.00214	0.001776	0.004444	0.003406	0.008139	0.007197	0.007719	0.001561	0.003636	0.008598	0.010029	0.007296	0.000627	0.000231	0.000907	0.000278	0.002634	0.001198	0.000325	0	0	0.000358	0.000261	0.000866
п	0.002393	7.98E-05	0.000138	2.59E-05	3.66E-05	0.002869	1.08E-05	3.66E-05	0.001705	0	0.000129	0.000996	3.02E-05	0.000263	0.01146	0.001332	0.006904	0.000207	0.00017	0.001035	6.47E-06	8.62E-06	8.62E-06	3.66E-05	1.94E-05	6.47E-06	0.000358	7.54E-05	2.16E-05	4.31E-06	0.000524
р	0.009266	0.000209	0.000649	0.000485	0.000453	0.006731	0.000422	0.000198	0.006143	2.16E-06	0.000586	0.000168	0.000584	0.001759	0.009389	0.000362	0.00014	0.000892	0.000886	0.003222	6.25E-05	0.000231	9.27E-05	0.000136	0.000222	6.9E-05	0.001789	0.000599	8.62E-06	0.000205	0.001073
С	0.00219	0.000222	0.001832	0.00016	0.000358	0.003505	0.000121	0.00011	0.002509	0	0.005371	0.002789	0.001104	0.001459	0.003304	0.002466	0.000373	0.001345	0.014357	0.000916	4.53E-05	0.000237	7.76E-05	0.00039	9.7E-05	0	0.000494	0.003434	7.11E-05	0.000302	0.004347
T	0.007999	0.000328	0.003822	0.000136	0.000522	0.007022	9.91E-05	0.000157	0.004789	0	0.000918	0.000317	0.001101	0.001938	0.011672	0.000688	0.0041	0.001714	0.000319	0.002746	2.59E-05	8.62E-05	0.00014	0.000272	5.6E-05	4.31E-05	0.001845	0.005266	9.91E-05	0.000151	0.000718
У	0.000246	0.0015	0.001332	0.001054	0.002242	0.000384	0.001584	0.000634	0.000668	0.000166	0.001716	0.002724	0.001481	0.001203	0.00045	0.001625	0.001231	0.002054	0.001933	0.000172	7.54E-05	0.000685	3.66E-05	0.001188	0.00083	0.000272	0	0	9.05E-05	0.001211	0.00016
ф	0.000291	4.31E-06	6.47E-06	2.16E-06	6.47E-06	0.000541	0	2.16E-06	0.000345	0	4.31E-06	8.84E-05	1.08E-05	1.51E-05	0.000364	4.31E-06	0.000106	2.59E-05	2.8E-05	0.000101	2.16E-05	0	0	0	0	0	1.29E-05	3.02E-05	0	4.31E-06	0
x	0.000963	0.000168	0.000535	0.000125	0.000175	0.000175	6.47E-05	9.7E-05	0.000612	0	0.000282	0.00033	0.000347	0.000489	0.002821	0.00047	0.0003	0.0005	0.000181	0.00031	2.37E-05	2.37E-05	1.94E-05	6.04E-05	5.17E-05	2.16E-06	0	2.16E-06	5.6E-05	0	2.8E-05
ц	0.001123	6.47E-06	0.000235	1.72E-05	1.94E-05	0.001211	0	1.29E-05	0.000931	0	0.000265	6.47E-06	2.37E-05	2.8E-05	0.000246	3.66E-05	1.72E-05	4.74E-05	1.72E-05	0.000325	0.000233	6.47E-06	2.16E-06	4.31E-06	2.16E-06	2.16E-06	0.000325	0	4.31E-06	0	0
ч	0.002625	6.04E-05	0.000153	1.51E-05	5.17E-05	0.003901	4.31E-06	7.54E-05	0.002033	0	0.000851	9.91E-05	3.45E-05	0.001136	0.000231	0.000194	9.91E-05	0.00017	0.002093	0.000726	2.16E-06	2.16E-06	0	4.53E-05	8.84E-05	2.16E-06	0	0.000254	1.51E-05	0	1.51E-05
ш	0.001207	4.31E-06	4.96E-05	6.47E-06	2.16E-06	0.002349	0	0	0.00192	0	0.000552	0.00061	1.72E-05	0.000463	0.000366	3.66E-05	4.31E-06	1.51E-05	0.000168	0.000394	0	1.72E-05	4.31E-06	4.31E-06	2.16E-06	0	0	0.000291	0	6.47E-06	2.16E-06
щ	0.000554	4.53E-05	1.94E-05	1.08E-05	2.16E-06	0.001444	2.16E-06	4.31E-06	0.000985	0	6.47E-06	0	2.16E-06	6.9E-05	6.47E-06	4.31E-06	2.16E-06	8.62E-06	6.47E-06	0.000172	2.16E-06	0	0	2.16E-06	0	2.16E-06	0	3.88E-05	0	0	0
ы	6.47E-05	0.00042	0.001597	0.000323	0.000431	0.001742	7.98E-05	0.000274	0.000558	0.002334	0.000513	0.002257	0.001791	0.000787	0.000291	0.000877	0.00044	0.00131	0.000935	0.000136	2.37E-05	0.001412	2.37E-05	0.000237	0.00055	2.59E-05	0	0	5.6E-05	2.16E-06	5.17E-05
ь	0.000185	0.00044	0.001073	0.000444	0.000616	0.000974	7.98E-05	0.000455	0.000752	0	0.001584	0.000123	0.000662	0.002464	0.000653	0.000996	0.000308	0.001722	0.000679	0.000216	6.9E-05	0.000153	0.000134	0.000394	0.000694	3.23E-05	0	0	0.000151	0.00047	0.001173
9	0	0	0	0	4.31E-06	0	0	2.16E-06	0	2.16E-06	0.000103	0.00017	9.48E-05	3.02E-05	2.16E-06	4.1E-05	7.76E-05	5.6E-05	0.001828	0	2.37E-05	2.37E-05	0	0	2.16E-06	0	0	0	0	0	0
ю	4.53E-05	0.000448	0.000276	9.48E-05	0.000494	3.66E-05	0.000101	0.000127	0.000185	6.47E-06	0.000399	9.27E-05	0.000228	0.00025	0.000131	0.000315	0.000188	0.000351	0.000591	5.6E-05	3.45E-05	3.88E-05	4.31E-05	0.000207	6.25E-05	0.000489	0	0	2.59E-05	4.1E-05	4.1E-05
Я	0.00017	0.000494	0.001543	0.000405	0.000935	0.000321	0.000317	0.000556	0.00072	0.000129	0.000938	0.001226	0.000793	0.002209	0.000623	0.001242	0.00047	0.00164	0.002097	0.000267	4.96E-05	0.000321	0.000106	0.000483	9.27E-05	0.000304	0	0	0.00011	0.00014	0.00016

Ентропія біграм, що перетинаються, в тексті без пробілів:

H.2 = 4.165314509471507

Надлишковість для біграм, що перетинаються, без пробілів:

R.2 = 0.15923507093600897

Частота біграм, що не перетинаються, в тексті без пробілів: bigram_table_not_cross_not_spaces.xlsx

	a	6	В	r	А	e	ж	3	и	й	к	л	м	н	0	п	р	c	T	у	ф	x	ц	ч	ш	щ	ы	ь	9	ю	я
a	0.000353	0.001599	0.005561	0.001405	0.003246	0.001436	0.001811	0.003746	0.001457	0.001035	0.005113	0.010346	0.004962	0.00635	0.001155	0.004281	0.004776	0.006509	0.007406	0.000707	0.0005	0.001681	0.000392	0.001448	0.001177	0.000297	0	0	0.000349	0.000862	0.002578
6	0.001112	2.16E-05	6.04E-05	4.31E-06	3.02E-05	0.002992	3.02E-05	1.29E-05	0.001185	0	0.000116	0.001168	4.31E-05	0.000328	0.002513	3.02E-05	0.001504	0.000194	2.16E-05	0.001164	4.31E-06	4.31E-05	2.59E-05	2.59E-05	4.31E-05	0.000194	0.003496	0.000466	3.02E-05	8.62E-05	0.000323
В	0.007514	0.000302	0.000427	0.000345	0.000836	0.006656	9.05E-05	0.00066	0.00472	0	0.00119	0.001108	0.000448	0.001651	0.007945	0.001039	0.001073	0.00335	0.000979	0.001043	5.17E-05	0.00019	0.000134	0.000241	0.000961	1.29E-05	0.003703	9.91E-05	0.000211	4.31E-06	0.000358
r	0.001759	3.02E-05	6.47E-05	8.62E-06	0.000862	0.000358	4.31E-06	3.45E-05	0.000789	0	0.000181	0.001272	2.16E-05	0.000392	0.008285	9.91E-05	0.00178	0.000108	2.16E-05	0.000819	0	0	0	3.02E-05	0	0	0	0	8.62E-06	0	4.31E-06
Д	0.005065	9.05E-05	0.00172	4.74E-05	0.000108	0.006173	0.000108	6.9E-05	0.002828	0	0.000487	0.000634	0.000151	0.002095	0.004651	0.000315	0.001026	0.000655	0.000198	0.00178	0	8.62E-05	0.000522	2.16E-05	8.19E-05	0	0.000672	0.000487	8.62E-06	3.02E-05	0.000461
e	0.000388	0.002194	0.004811	0.003263	0.004505	0.002354	0.00119	0.00216	0.001112	0.002599	0.002556	0.007859	0.004695	0.009777	0.001267	0.003022	0.007734	0.007497	0.006341	0.000767	0.000147	0.00097	0.000651	0.001647	0.001009	0.000789	0	0	0.00025	0.000112	0.00041
ж	0.001673	2.59E-05	2.16E-05	1.29E-05	0.000897	0.003781	4.31E-06	2.16E-05	0.001703	0	0.00025	3.02E-05	0	0.001095	9.91E-05	3.45E-05	1.72E-05	3.45E-05	4.31E-06	0.00028	8.62E-06	0	4.31E-06	3.88E-05	0	0	0	2.59E-05	4.31E-06	0	4.31E-06
3	0.006466	0.000254	0.001216	0.000289	0.001108	0.00072	9.91E-05	7.76E-05	0.000621	0	0.00025	0.000228	0.000405	0.001686	0.000841	0.000254	0.000306	0.000168	0.000116	0.000448	1.29E-05	4.31E-06	3.45E-05	0.000116	2.16E-05	0	0.00041	0.000168	8.62E-06	4.31E-06	0.000289
И	0.000504	0.001285	0.004703	0.001276	0.002729	0.002543	0.000457	0.003311	0.00194	0.001776	0.004358	0.006294	0.003237	0.006531	0.002104	0.003755	0.002065	0.005833	0.006445	0.000526	0.000228	0.00191	0.001207	0.003388	0.000711	0.000427	4.31E-06	0	0.000203	0.000362	0.001375
й	0.000185	0.000414	0.000711	0.000392	0.000845	0.000112	0.00019	0.000319	0.000625	0	0.000987	0.000306	0.000491	0.001125	0.000526	0.00097	0.000315	0.001617	0.000918	0.000181	7.33E-05	7.76E-05	0.000138	0.000535	0.000181	2.59E-05	0	0	8.19E-05	1.72E-05	9.48E-05
К	0.007583	0.000203	0.000897	0.000125	0.000237	0.000957	0.000108	0.00019	0.00488	0	0.000241	0.001048	0.000172	0.000897	0.010497	0.000461	0.002574	0.000767	0.001173	0.002285	4.31E-06	3.88E-05	0.000224	0.000172	3.45E-05	0	0	0	7.76E-05	4.31E-06	5.6E-05
Л	0.006001	0.000366	0.001117	0.000466	0.000341	0.005695	0.000323	0.000267	0.009958	4.31E-06	0.001086	0.000414	0.000194	0.001185	0.008277	0.000888	0.000207	0.00266	0.000448	0.001742	1.72E-05	5.6E-05	1.72E-05	0.000466	7.76E-05	8.62E-06	0.000741	0.004919	0.000103	0.000979	0.001983
м	0.00447	0.000655	0.000948	0.000388	0.000401	0.004044	0.000147	0.000246	0.004302	0	0.00069	0.000341	0.0005	0.001888	0.004608	0.001272	0.000332	0.000974	0.00028	0.002216	8.19E-05	6.47E-05	4.74E-05	0.000207	0.000129	4.31E-06	0.00097	0.000129	8.62E-05	4.31E-06	0.000569
н	0.010971	0.000185	0.000366	0.00025	0.001022	0.008816	0.000168	0.000172	0.008135	0	0.000789	3.88E-05	0.000116	0.00344	0.011406	0.000448	0.000147	0.00131	0.00116	0.003337	3.45E-05	4.31E-05	0.000716	0.000414	3.88E-05	9.48E-05	0.004884	0.001272	3.45E-05	0.000138	0.00122
0	0.00025	0.00494	0.011467	0.005143	0.006953	0.002216	0.002647	0.002229	0.001681	0.004393	0.003457	0.008083	0.007311	0.007712	0.001539	0.003621	0.008544	0.009937	0.007212	0.000664	0.000198	0.000905	0.000276	0.002681	0.00125	0.000379	0	0	0.000362	0.000272	0.000832
п	0.002401	7.76E-05	0.000147	3.02E-05	3.45E-05	0.002759	4.31E-06	2.59E-05	0.001746	0	9.48E-05	0.000991	4.31E-05	0.000302	0.011402	0.00128	0.007014	0.000185	0.000203	0.001004	8.62E-06	1.29E-05	0	3.88E-05	3.02E-05	0	0.000358	7.76E-05	1.29E-05	0	0.000513
p	0.009424	0.000177	0.000616	0.000474	0.000435	0.006656	0.00044	0.00019	0.006044	4.31E-06	0.000582	0.000134	0.000591	0.001711	0.009303	0.000345	0.000134	0.000875	0.000914	0.003306	5.6E-05	0.000216	0.000121	0.000125	0.000267	6.9E-05	0.001841	0.000612	0	0.000181	0.001099
c	0.002263	0.000228	0.001841	0.000151	0.000397	0.003453	0.000142	9.48E-05	0.002384	0	0.005397	0.002802	0.001017	0.00144	0.003513	0.002436	0.000332	0.001375	0.014726	0.00091	4.31E-05	0.000254	7.33E-05	0.000427	0.000108	0	0.000513	0.003311	8.62E-05	0.00028	0.004427
T	0.007742	0.000319	0.003975	0.000147	0.00047	0.006811	9.91E-05	0.000172	0.004712	0	0.000901	0.000319	0.001117	0.001884	0.011596	0.000655	0.004078	0.001673	0.000358	0.002703	2.59E-05	9.91E-05	0.000134	0.00028	7.76E-05	4.74E-05	0.001836	0.005466	0.000116	0.000138	0.000716
у	0.000241	0.001492	0.001302	0.001039	0.002311	0.000388	0.001638	0.000664	0.000746	0.000164	0.001755	0.002776	0.001509	0.00119	0.000431	0.001599	0.001177	0.002043	0.001966	0.000168	7.33E-05	0.000724	3.02E-05	0.001203	0.000785	0.000254	0	0	7.76E-05	0.001207	0.000151
ф	0.000276	4.31E-06	8.62E-06	0	4.31E-06	0.000522	0	0	0.000306	0	0	8.62E-05	1.29E-05	1.29E-05	0.000341	4.31E-06	0.000116	1.29E-05	2.59E-05	9.91E-05	2.16E-05	0	0	0	0	0	1.29E-05	3.88E-05	0	0	0
x	0.00103	0.000147	0.000552	0.000108	0.000177	0.000147	5.6E-05	8.62E-05	0.000629	0	0.000254	0.000293	0.000341	0.000487	0.002763	0.000491	0.000293	0.0005	0.00016	0.000272	2.59E-05	2.59E-05	2.16E-05	6.47E-05	5.6E-05	0	0	4.31E-06	5.17E-05	0	3.02E-05
ц	0.001142	0	0.000216	1.29E-05	2.59E-05	0.001203	0	1.72E-05	0.000923	0	0.000302	4.31E-06	3.02E-05	3.45E-05	0.000224	4.31E-05	2.16E-05	3.88E-05	2.16E-05	0.000366	0.000263	8.62E-06	0	4.31E-06	0	0	0.000328	0	4.31E-06	0	0
ч	0.00266	5.6E-05	0.000138	1.72E-05	4.31E-05	0.003824	8.62E-06	6.47E-05	0.002043	0	0.000918	0.000103	3.88E-05	0.001078	0.000241	0.000203	7.76E-05	0.000142	0.002082	0.000729	4.31E-06	0	0	4.31E-05	9.91E-05	4.31E-06	0	0.000267	1.72E-05	0	2.16E-05
ш	0.001177	8.62E-06	3.88E-05	0	4.31E-06	0.002311	0	0	0.001867	0	0.00056	0.000586	2.16E-05	0.000491	0.000306	3.02E-05	4.31E-06	2.59E-05	0.000168	0.000422	0	2.16E-05	4.31E-06	8.62E-06	4.31E-06	0	0	0.000306	0	1.29E-05	0
щ	0.000539	3.88E-05	2.16E-05	4.31E-06	0	0.001358	0	0	0.000979	0	8.62E-06	0	0	9.05E-05	1.29E-05	0	4.31E-06	8.62E-06	4.31E-06	0.000155	4.31E-06	0	0	0	0	4.31E-06	0	2.16E-05	0	0	0
ы	3.88E-05	0.000435	0.001573	0.000358	0.00041	0.00175	6.47E-05	0.000293	0.000578	0.002367	0.000453	0.002125	0.001793	0.000754	0.00031	0.000823	0.000418	0.001336	0.000935	0.000116	2.59E-05	0.001392	3.02E-05	0.000246	0.000552	3.45E-05	0	0	4.74E-05	4.31E-06	5.17E-05
ь	0.000185	0.000414	0.001142	0.000414	0.000591	0.000888	8.19E-05	0.000526	0.000785	0	0.001565	0.000134	0.000655	0.002513	0.00066	0.000974	0.000297	0.001673	0.000694	0.000241	8.62E-05	0.000168	0.000116	0.000366	0.000746	3.88E-05	0	0	0.000134	0.000466	0.001168
9	0	0	0	0	4.31E-06	0	0	0	0	4.31E-06	8.62E-05	0.000177	9.05E-05	3.02E-05	0	3.45E-05	9.05E-05	3.88E-05	0.001806	0	2.59E-05	2.59E-05	0	0	4.31E-06	0	0	0	0	0	0
ю	2.59E-05	0.00044	0.000272	9.05E-05	0.000509	4.31E-05	0.000116	0.000116	0.000198	8.62E-06	0.000384	0.000103	0.000228	0.000198	0.000134	0.000319	0.000211	0.000379	0.000591	5.17E-05	4.74E-05	3.45E-05	3.45E-05	0.000254	8.19E-05	0.000496	0	0	3.45E-05	3.02E-05	4.31E-05
я	0.000172	0.000513	0.001457	0.000414	0.000974	0.000306	0.000276	0.000543	0.000737	0.000116	0.000914	0.001302	0.000763	0.002112	0.000591	0.001254	0.000461	0.001625	0.00213	0.000289	5.17E-05	0.000306	0.000112	0.000539	8.19E-05	0.000345	0	0	0.000103	0.000129	0.000164

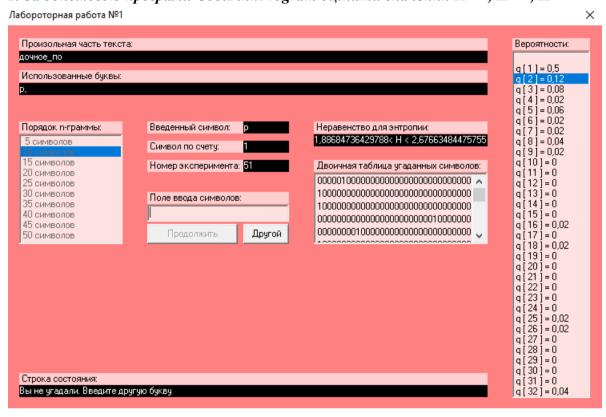
Ентропія біграм, що не перетинаються, в тексті без пробілів:

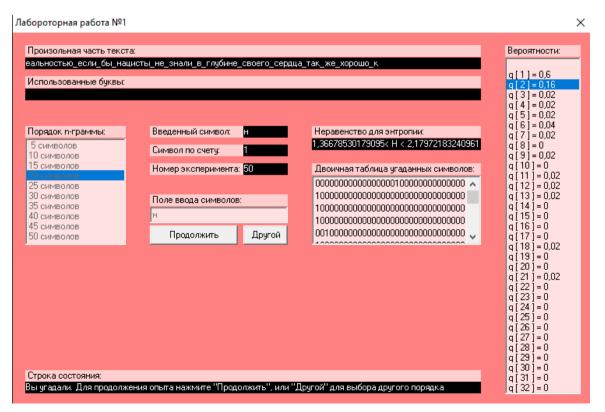
H.2 = 4.162903243557497

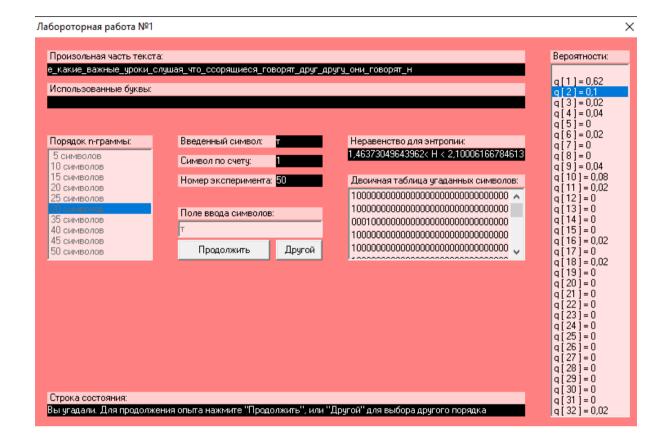
Надлишковість для біграм, що не перетинаються, без пробілів:

R.2 = 0.15972178275825843

2. За допомогою програми CoolPinkProgram оцінити значення $H^{(10)}$, $H^{(20)}$, $H^{(30)}$







Маємо наступні значення:

$$1.88684736429788 < H^{(10)} < 2.67663484475755$$

$$1.36678530179095 < H^{(20)} < 2.17972183240961$$

$$1.46373049643962 < H^{(30)} < 2.10006166784613$$

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

Hадлишковість джерела відкритого тексту (мови) дорівнює $R=1-\frac{H_{\infty}}{H_0}$ $H_0=\log_2(32)=5$ (32 букви алфавіту, враховуючи пробіл)

Для H⁽¹⁰⁾:

$$R = 1 - \frac{1.88684736429788}{5} = 0.622630527140424$$

$$R = 1 - \frac{2.67663484475755}{5} = 0.46467303104849$$

Для H⁽²⁰⁾:

$$R = 1 - \frac{1.36678530179095}{5} = 0.72664293964181$$

$$R = 1 - \frac{2.17972183240961}{5} = 0.564055633518078$$

Для H⁽³⁰⁾:

$$R = 1 - \frac{1.46373049643962}{5} = 0.707253900712076$$

$$R = 1 - \frac{2.10006166784613}{5} = 0.579987666430774$$

Для підсумку маємо:

$$0.46467303104849 < R^{(10)} < 0.622630527140424$$

$$0.564055633518078 < R^{(20)} < 0.72664293964181$$

$$0.579987666430774 < R^{(30)} < 0.707253900712076$$

Труднощі при виконанні роботи:

Найскладнішим моментом при виконанні лабораторної роботи було побороти прокрастинацію та всістись за виконання завдань

На початку, основним завданням було відфільтрувати текст, ми використали Regular Expressions для швидкої обробки. Для аналізу тексту беруться тексти із пробілами та без них, проте ця проблема була вирішена доволі швидко шляхом модифікації функцій і додавання логічного параметру spaces. Так само були модифіковані функції letter_frequency() та bigram_frequency(), в останній також додано параметр cross, що відповідає за те, чи є біграма тою, у якій пари букв перетинаються, чи ні. Шляхом модифікації було скорочено код

Також, проблемою було правильно вивести та зберегти таблицю частот біграм. Вирішено проблему було за рахунок створення датафрейму pandas та його імпорту у Excel таблицю

Варто зазначити, що ми створили 2 функції для літер та біграм, які використовували для різних типів біграм та пробілів тексту, що робить код більше компактним. Окрім частот, вони також виводять підраховні ентропії та надлишковості для цих типів

Висновки:

Під час виконання лабораторної роботи ми ознайомилимь із наступними термінами:

Ентропія - міра невизначеності, показує кількість необхідної інформації для однозначного опису ансамблю (пара з множини символів та множини ймовірностей). Вирахувавши частоту появу символів алфавіту можна визначити ентропію тексту, тобто скільки інформації містить текст

Сукупна ентропія говорить нам, скільки інформації містять два ансамблі (із урахуванням взаємних залежностей)

Умовна ентропія говорить нам, скільки інформації залишиться в ансамблі X, якщо поведінку ансамблю Y буде однозначно визначено

На початку роботи для аналізу та підрахунку частоти та ентропії букв і біграм було обрано текст "Двенадцать стульев". Було написано програму на мові Python3 $razik_bolgov.py$, що підраховує частоту букв і частоту біграм в тексті, а також ентропії H_1 та H_2 з надлишковістю

Було встановлено, що у російській мові (без пробілу) найчастіше трапляються літери "о", "а", "е", "и", "н", "т", "с", "л", "в", "р". Також було встановлено частоту різних біграм. Ці знання допомогли нам при експерименті з підрахунку надлишковості російської мови - при роботі із CoolPinkProgram, якщо ми не могли визначити наступну літеру із попередніх то вводили літери методом брутфорсу (від самої популярної), в більшості таких випадків вгадування літери було доволі швидким

В ході виконання лабораторної роботи при роботі із CoolPinkProgram ми також стикнулись із поняттям **умовної ентропії джерела**, що визначає, скільки інформації про наступний символ ми матимемо із значень попередніх. Було оцінено умовну ентропію при n=10, 20, 30. Завдяки чому було підраховано **надлишковість** російської **мови** (величину можливого ущільнення тексту деякою схемою кодування символів без втрати його змісту)