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:
Красноок Юлія та Водяник Дмитро

Київ - 2025

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.

Постановка задачі

1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку $H1$ та $H2$ за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення $H1$ та $H2$ на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення $H1$ та $H2$ на тому ж тексті, в якому вилучено всі пробіли.

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

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

Частоти появи окремих літер у тексті російською мовою(без пробілів)

Symbol	Frequency
о	0.112613
е	0.082056
а	0.078744
и	0.071659
н	0.062486
т	0.057288
с	0.052433
в	0.049190
л	0.046733
р	0.043041
к	0.038846
д	0.030325
у	0.030094
м	0.028734
п	0.027146
ы	0.021021
ь	0.019702
я	0.018956
г	0.018717
б	0.018668

з	0.017342
ч	0.015776
ж	0.010865
х	0.010771
й	0.010448
ш	0.009407
ю	0.006566
ц	0.003941
щ	0.002905
э	0.001767
ф	0.001202
ё	0.000308
ъ	0.000255

H1 без пробілів:

Ентропія: 4.471536

H2 без пробілів з перетином:

Ентропія: 3.813652

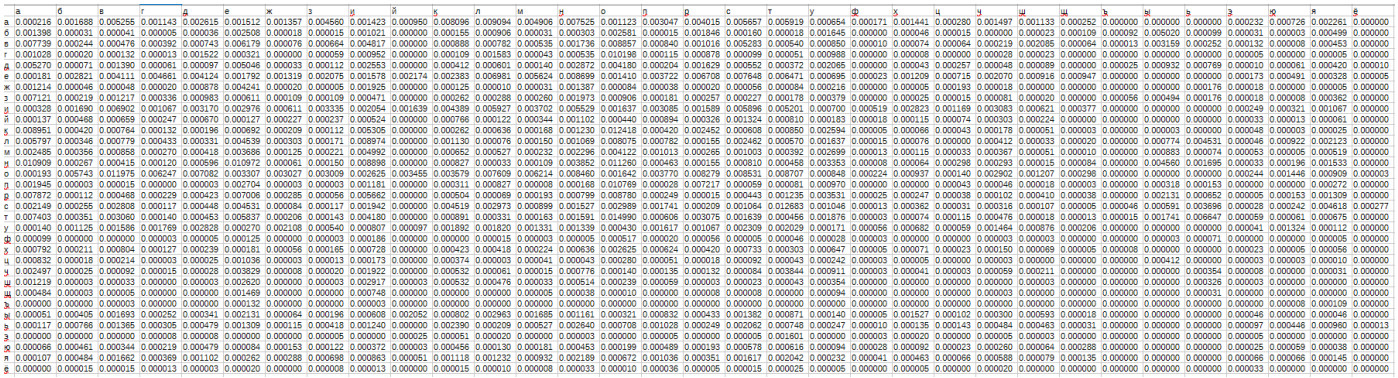
H2 без пробілів без перетину:

Ентропія: 3.806673

Частота біграм без пробілів без перетину

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я	ё
а	0.000178	0.001578	0.005295	0.001166	0.002490	0.001553	0.001390	0.004358	0.001431	0.001018	0.007887	0.009231	0.004969	0.007225	0.001156	0.003075	0.003977	0.005616	0.005876	0.000703	0.000214	0.001497	0.000260	0.001502	0.001141	0.000265	0.000000	0.000000	0.000000	0.000229	0.000723	0.002306	0.000000
б	0.001477	0.000031	0.000056	0.000005	0.000046	0.002581	0.000025	0.000020	0.000998	0.000000	0.000168	0.000927	0.000025	0.000311	0.002643	0.000005	0.001731	0.000183	0.000025	0.001634	0.000000	0.000031	0.000015	0.000000	0.000031	0.000092	0.000107	0.004985	0.000087	0.000015	0.000000	0.000474	0.000000
в	0.007754	0.000270	0.000428	0.000367	0.000728	0.005988	0.000056	0.000631	0.004827	0.000000	0.000918	0.000789	0.000540	0.001787	0.009653	0.000820	0.001039	0.005392	0.000621	0.000815	0.000015	0.000061	0.000051	0.000229	0.002159	0.000081	0.000015	0.003050	0.000229	0.000102	0.000010	0.000458	0.000000
г	0.001018	0.000025	0.000112	0.000005	0.001517	0.000336	0.000000	0.000051	0.001013	0.000000	0.000092	0.001563	0.000051	0.000443	0.010260	0.000132	0.000901	0.000081	0.000056	0.001090	0.000000	0.000000	0.000000	0.000020	0.000015	0.000000	0.000000	0.000000	0.000000	0.000005	0.000000	0.000005	0.000000
д	0.005204	0.000071	0.001446	0.000061	0.000087	0.005066	0.000036	0.000122	0.002658	0.000000	0.000412	0.000555	0.000132	0.002917	0.004267	0.000193	0.001716	0.000468	0.000392	0.001955	0.000000	0.000031	0.000239	0.000066	0.000087	0.000000	0.000036	0.000932	0.000754	0.000020	0.000081	0.000474	0.000010
е	0.000183	0.003019	0.004002	0.004547	0.004175	0.001757	0.001273	0.001889	0.001456	0.002317	0.002500	0.007087	0.005499	0.008803	0.001309	0.003829	0.006645	0.007434	0.006507	0.000784	0.000015	0.001171	0.000687	0.002062	0.000866	0.000978	0.000000	0.000000	0.000000	0.000158	0.000499	0.000387	0.000005
ж	0.001227	0.000041	0.000046	0.000025	0.000896	0.004358	0.000020	0.000005	0.001920	0.000000	0.000117	0.000010	0.000020	0.001380	0.000056	0.000036	0.000010	0.000066	0.000081	0.000239	0.000000	0.000005	0.000178	0.000015	0.000000	0.000000	0.000000	0.000000	0.000178	0.000015	0.000000	0.000005	0.000000
з	0.007454	0.000204	0.001258	0.000351	0.001090	0.000621	0.000117	0.000092	0.000509	0.000000	0.000219	0.000275	0.000270	0.002057	0.001008	0.000193	0.000255	0.000265	0.000168	0.000382	0.000000	0.000010	0.000015	0.000092	0.000015	0.000000	0.000051	0.000504	0.002099	0.000020	0.000005	0.000356	0.000000
и	0.000300	0.001808	0.006925	0.001110	0.003243	0.003055	0.000611	0.003447	0.002067	0.001675	0.004333	0.005753	0.003605	0.005591	0.001589	0.003182	0.001619	0.006059	0.005316	0.000799	0.000499	0.002907	0.001232	0.003101	0.000601	0.000336	0.000000	0.000000	0.000000	0.000249	0.000321	0.001166	0.000000
й	0.000137	0.000479	0.000601	0.000265	0.000616	0.000132	0.000239	0.000275	0.000545	0.000000	0.000713	0.000137	0.000295	0.001135	0.000387	0.000789	0.000326	0.001319	0.000769	0.000244	0.000020	0.000137	0.000076	0.000270	0.000234	0.000000	0.000000	0.000000	0.000000	0.000031	0.000015	0.000056	0.000000
к	0.009094	0.000418	0.000713	0.000163	0.000173	0.000606	0.000224	0.000117	0.005158	0.000000	0.000260	0.000692	0.000178	0.001125	0.012525	0.000382	0.002561	0.000631	0.000860	0.002668	0.000005	0.000066	0.000041	0.000153	0.000046	0.000005	0.000000	0.000000	0.000000	0.000041	0.000005	0.000020	0.000000
л	0.005804	0.000331	0.000733	0.000438	0.000305	0.004511	0.000336	0.000188	0.008900	0.000000	0.001232	0.000071	0.000188	0.001034	0.008157	0.000820	0.000148	0.002622	0.000494	0.001706	0.000010	0.000076	0.000000	0.000407	0.000041	0.000025	0.000000	0.000789	0.004659	0.000048	0.008896	0.002088	0.000000
м	0.002592	0.000346	0.000881	0.000270	0.000407	0.003768	0.000102	0.000219	0.004995	0.000000	0.000540	0.000550	0.000229	0.002250	0.004358	0.000932	0.000290	0.000906	0.000392	0.002760	0.000015	0.000081	0.000020	0.000392	0.000061	0.000010	0.000000	0.000830	0.000076	0.000056	0.000000	0.000524	0.000000
н	0.011003	0.000249	0.000458	0.000102	0.000601	0.010947	0.000056	0.000097	0.008564	0.000000	0.000891	0.000046	0.000112	0.003865	0.011217	0.000428	0.000153	0.000840	0.000519	0.003452	0.000005	0.000061	0.000336	0.000285	0.000015	0.000081	0.000000	0.004297	0.001670	0.000036	0.000219	0.001492	0.000000
о	0.000224	0.005687	0.011975	0.006196	0.006975	0.003360	0.003009	0.002714	0.002551	0.003360	0.003651	0.007255	0.006232	0.006747	0.001634	0.003895	0.008045	0.008421	0.008651	0.000860	0.000239	0.000671	0.000137	0.002933	0.001191	0.000270	0.000000	0.000000	0.000000	0.000270	0.001436	0.000916	0.000005
п	0.001930	0.000000	0.000005	0.000000	0.000000	0.002755	0.000005	0.000000	0.001110	0.000000	0.000280	0.000774	0.000010	0.000168	0.019942	0.000020	0.007286	0.000051	0.000081	0.000911	0.000000	0.000000	0.000025	0.000051	0.000031	0.000000	0.000000	0.000326	0.000132	0.000000	0.000000	0.000249	0.000000
р	0.008029	0.000132	0.000468	0.000219	0.000438	0.007189	0.000295	0.000051	0.005585	0.000000	0.000494	0.000076	0.000178	0.000830	0.008844	0.002255	0.000020	0.000438	0.001283	0.003457	0.000020	0.000234	0.000025	0.000102	0.000372	0.000015	0.000000	0.002118	0.000621	0.000000	0.000163	0.001247	0.000000
с	0.002149	0.000209	0.002795	0.000092	0.000443	0.004470	0.000081	0.000132	0.001869	0.000000	0.000414	0.003080	0.000876	0.001548	0.002912	0.001614	0.000214	0.001125	0.013162	0.001100	0.000025	0.000295	0.000015	0.000311	0.000117	0.000000	0.000051	0.000550	0.003666	0.000036	0.000244	0.004496	0.000275
т	0.007098	0.000356	0.003040	0.000183	0.000479	0.005718	0.000234	0.000148	0.004185	0.000000	0.000804	0.000326	0.000183	0.001619	0.014852	0.000636	0.003091	0.001639	0.000438	0.001843	0.000000	0.000092	0.000132	0.000443	0.000025	0.000005	0.000010	0.001767	0.008568	0.000061	0.000051	0.000708	0.000000
у	0.000153	0.001105	0.001548	0.001808	0.002760	0.000260	0.000252	0.000614	0.000779	0.000112	0.001894	0.001731	0.001334	0.001329	0.000392	0.001543	0.001049	0.002408	0.001955	0.000137	0.000046	0.000698	0.000041	0.001421	0.000855	0.000209	0.000000	0.000000	0.000000	0.000051	0.001318	0.000087	0.000000
ф	0.000122	0.000000	0.000000	0.000000	0.000005	0.000137	0.000000	0.000005	0.000158	0.000000	0.000000	0.000010	0.000000	0.000005	0.000524	0.000020	0.000046	0.000000	0.000041	0.000010	0.000000	0.000000	0.000000	0.000000	0.000005	0.000000	0.000000	0.000005	0.000061	0.000000	0.000000	0.000000	0.000000
х	0.000840	0.000168	0.000825	0.000087	0.000209	0.000163	0.000046	0.000168	0.000677	0.000000	0.000428	0.000448	0.000229	0.000621	0.002811	0.000682	0.000412	0.000682	0.000249	0.000687	0.000010	0.000087	0.000015	0.000168	0.000066	0.000005	0.000010	0.000000	0.000000	0.000025	0.000005	0.000041	0.000000
ц	0.000779	0.000015	0.000229	0.000005	0.000015	0.001095	0.000005	0.000010	0.000183	0.000000	0.000418	0.000005	0.000041	0.000071	0.000260	0.000051	0.000020	0.000087	0.000025	0.000224	0.000005	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000423	0.000000	0.000005	0.000005	0.000005
ч	0.002429	0.000025	0.000097	0.000010	0.000015	0.003691	0.000000	0.000010	0.001909	0.000000	0.000550	0.000041	0.000010	0.000845	0.000158	0.000148	0.000102	0.000087	0.003885	0.000937	0.000005	0.000046	0.000005	0.000061	0.000239	0.000000	0.000000	0.000000	0.000321	0.000005	0.000000	0.000020	0.000000
ш	0.001135	0.000005	0.000031	0.000000	0.000005	0.002561	0.000000	0.000000	0.002877	0.000000	0.000670	0.000604	0.000031	0.000519	0.000280	0.000061	0.000005	0.000010	0.000051	0.003367	0.000000	0.000000	0.000000	0.000000	0.000005	0.000000	0.000000	0.000000	0.000336	0.000000	0.000000	0.000000	0.000000
щ	0.000530	0.000000	0.000010	0.000000	0.000000	0.001471	0.000000	0.000000	0.000799	0.000000	0.000000	0.000000	0.000005	0.000025	0.000015	0.000000	0.000010	0.000010	0.000000	0.000112	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000041	0.000000	0.000000	0.000000	0.000000
ъ	0.000000	0.000000	0.000005	0.000000	0.000000	0.000112	0.000000	0.000000	0.000005	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000005	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000005	0.000102	0.000000
ы	0.000036	0.000372	0.001690	0.000249	0.000351	0.002240	0.000056	0.000168	0.000601	0.002108	0.000855	0.003009	0.001782	0.001222	0.000336	0.000866	0.000392	0.001390	0.000840	0.000173	0.000010	0.001482	0.000102	0.000372	0.000652	0.000025	0.000000	0.000000	0.000000	0.000051	0.000000	0.000036	0.000000
ь	0.000137	0.000784	0.001309	0.000321	0.000509	0.001375	0.000107	0.000397	0.001273	0.000000	0.002368	0.000219	0.000624	0.002663	0.000672	0.001049	0.000244	0.002088	0.000748	0.000239	0.000010	0.000137	0.000158	0.000499	0.000433	0.000015	0.000000	0.000000	0.000000	0.000092	0.000504	0.000952	0.000000
э	0.000000	0.000000	0.000010	0.000005	0.000000	0.000000	0.000000	0.000000	0.000000	0.000025	0.000061	0.000020	0.000000	0.000005	0.000000	0.000005	0.000000	0.000005	0.001629	0.000000	0.000005	0.000015	0.000000	0.000000	0.000005	0.000000	0.000000	0.000000	0.000000	0.000005	0.000000	0.000000	0.000000
ю	0.000041	0.000418	0.000346	0.000219	0.000540	0.000097	0.000158	0.000107	0.000367	0.000000	0.000494	0.000107	0.000158	0.000443	0.000204	0.000504	0.000183	0.000570	0.000586	0.000066	0.000041	0.000092	0.000015	0.000295	0.000061	0.000260	0.000000	0.000000	0.000000	0.000020	0.000066	0.000031	0.000000
я	0.000122	0.000489	0.001752	0.000372	0.001039	0.000295	0.000244	0.000672	0.000830	0.000036	0.001181	0.001125	0.000916	0.002256	0.000682	0.000947	0.000382	0.001604	0.002138	0.000265	0.000031	0.000509	0.000076	0.000631	0.000092	0.000102	0.000000	0.000000	0.000000	0.000056	0.000071	0.000148	0.000000
ё	0.000000	0.000005	0.000020	0.000015	0.000000	0.000020	0.000000	0.000010	0.000015	0.000000	0.000020	0.000010	0.000010	0.000036	0.000010	0.000036	0.000010	0.000020	0.000020	0.000000	0.000000	0.000005	0.000000	0.000020	0.000000	0.000000	0.000000	0.000000	0.000000	0.000036	0.000000	0.000000	0.000000

Частота біграм без пробілів з перетином



Частоти появи окремих літер у тексті російською мовою(з пробілами)

Symbol	Frequency
	0.156851
о	0.094949
е	0.069185
а	0.066392
и	0.060419
н	0.052685
т	0.048302
с	0.044209
в	0.041474
л	0.039403
р	0.036290
к	0.032753
д	0.025569
у	0.025373
м	0.024227
п	0.022888
ы	0.017723
ь	0.016612
я	0.015983
г	0.015781
б	0.015740
з	0.014622
ч	0.013302
ж	0.009161
х	0.009082
й	0.008809

ш	0.007931
ю	0.005536
ц	0.003323
щ	0.002449
э	0.001490
ф	0.001013
ё	0.000260
ъ	0.000215

H1 з пробілами:

Ентропія: 4.396893

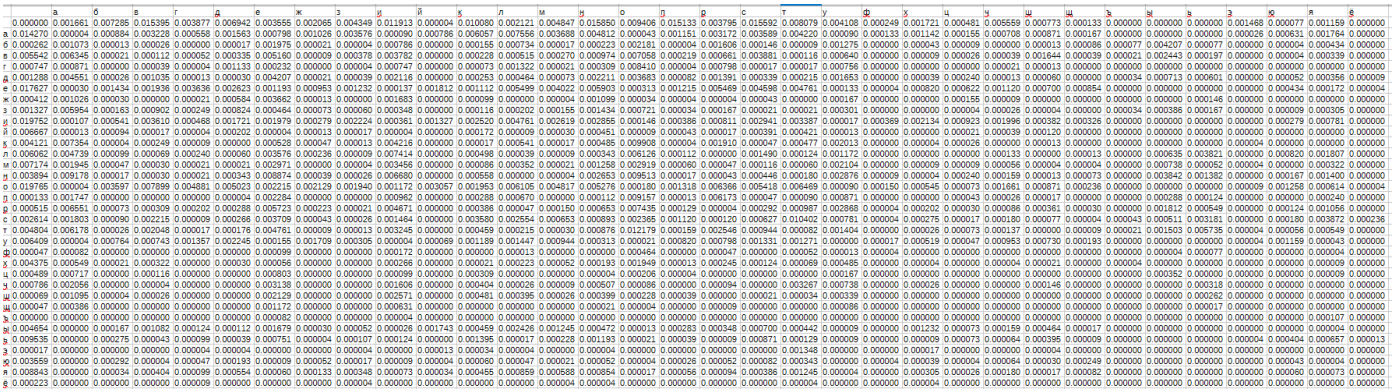
H2 з пробілами з перетином:

Ентропія: 3.579412

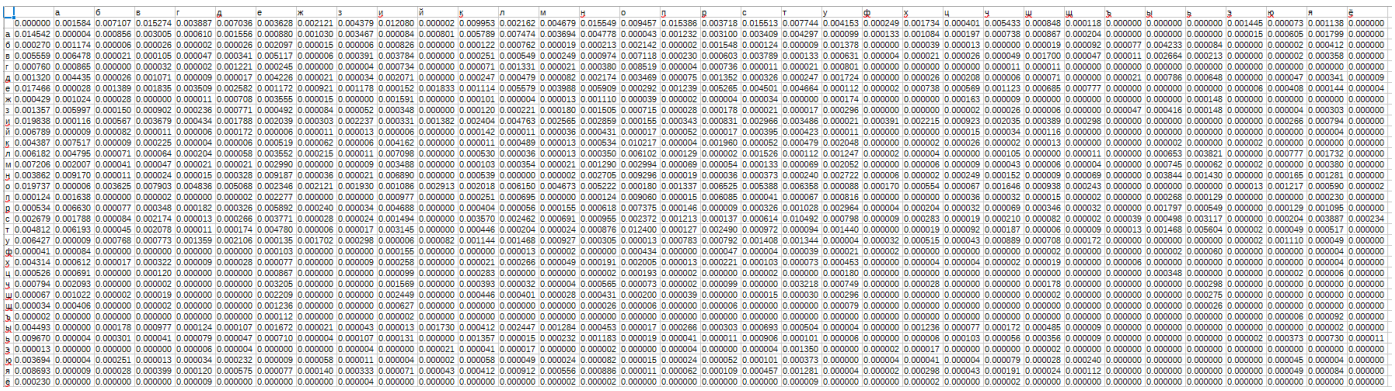
H2 з пробілами без перетину:

Ентропія: 3.581915

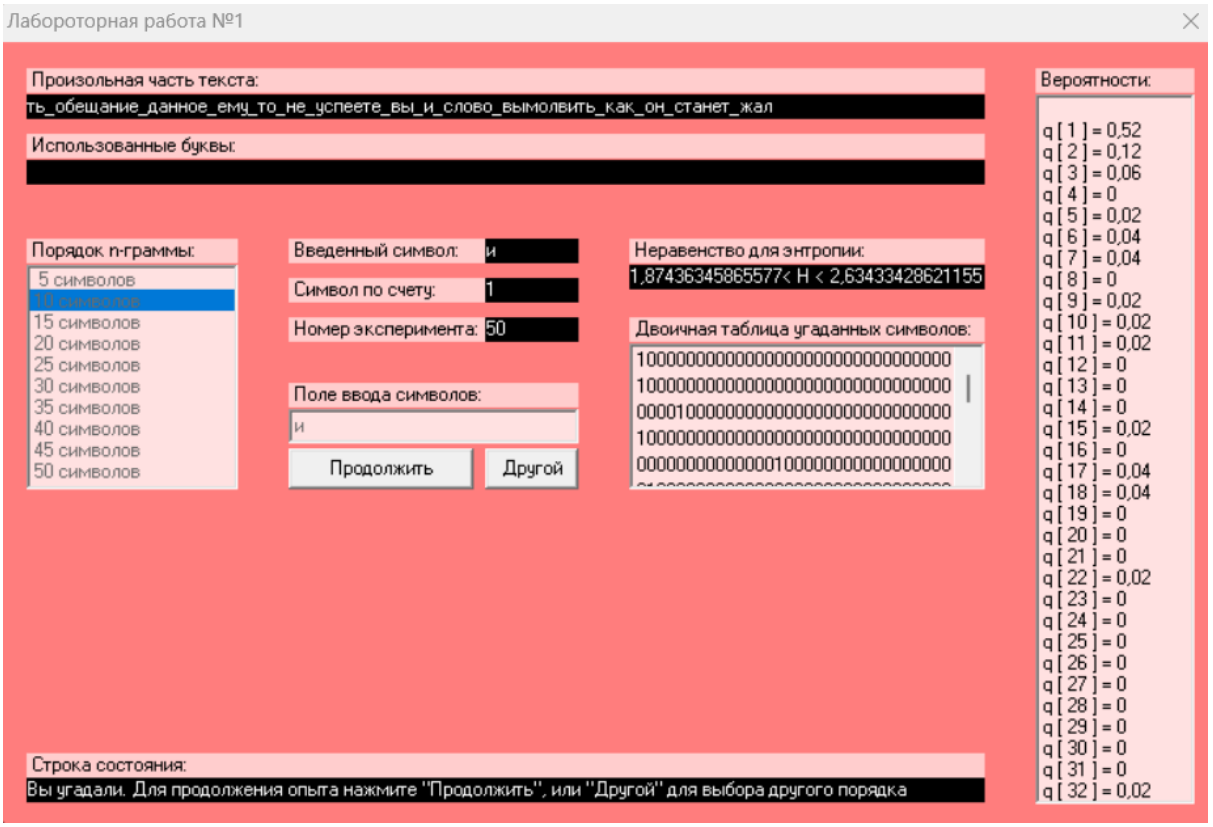
Частота біграм з пробілами без перетину



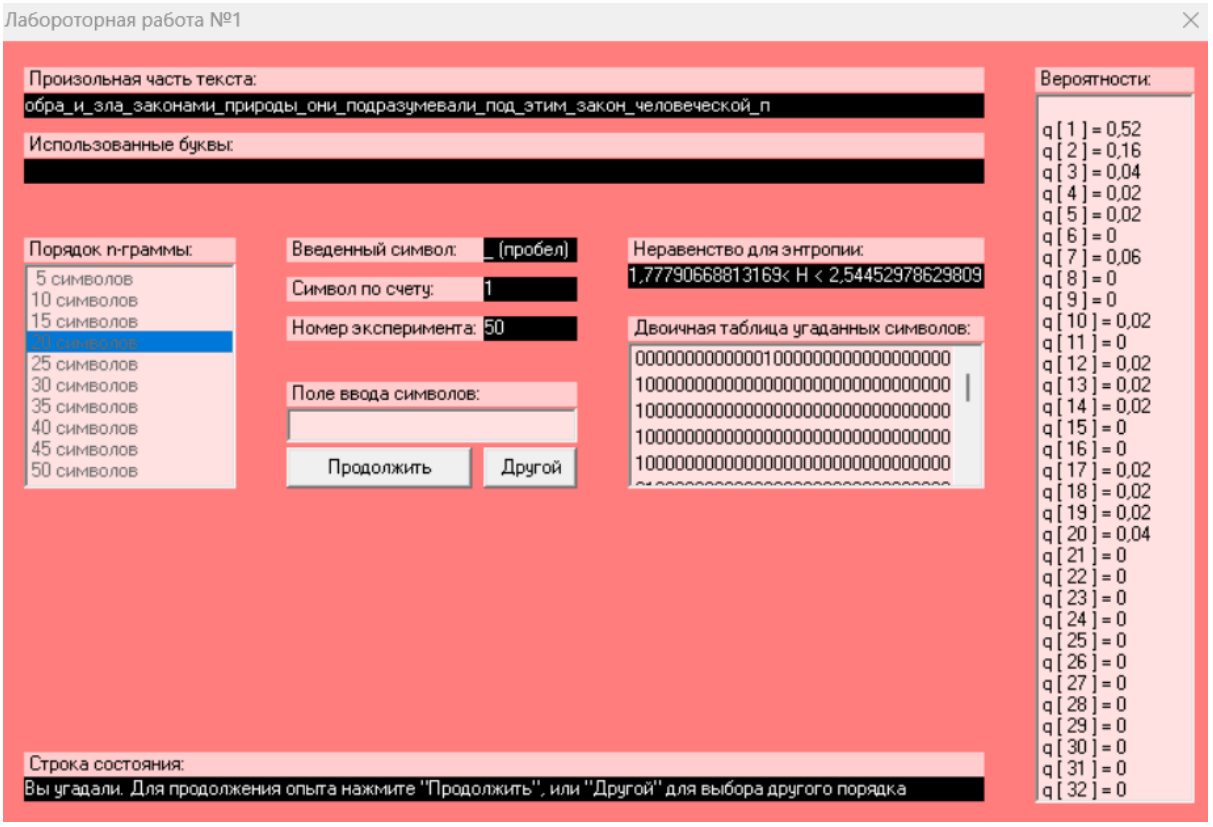
Частота біграм з пробілами з перетином



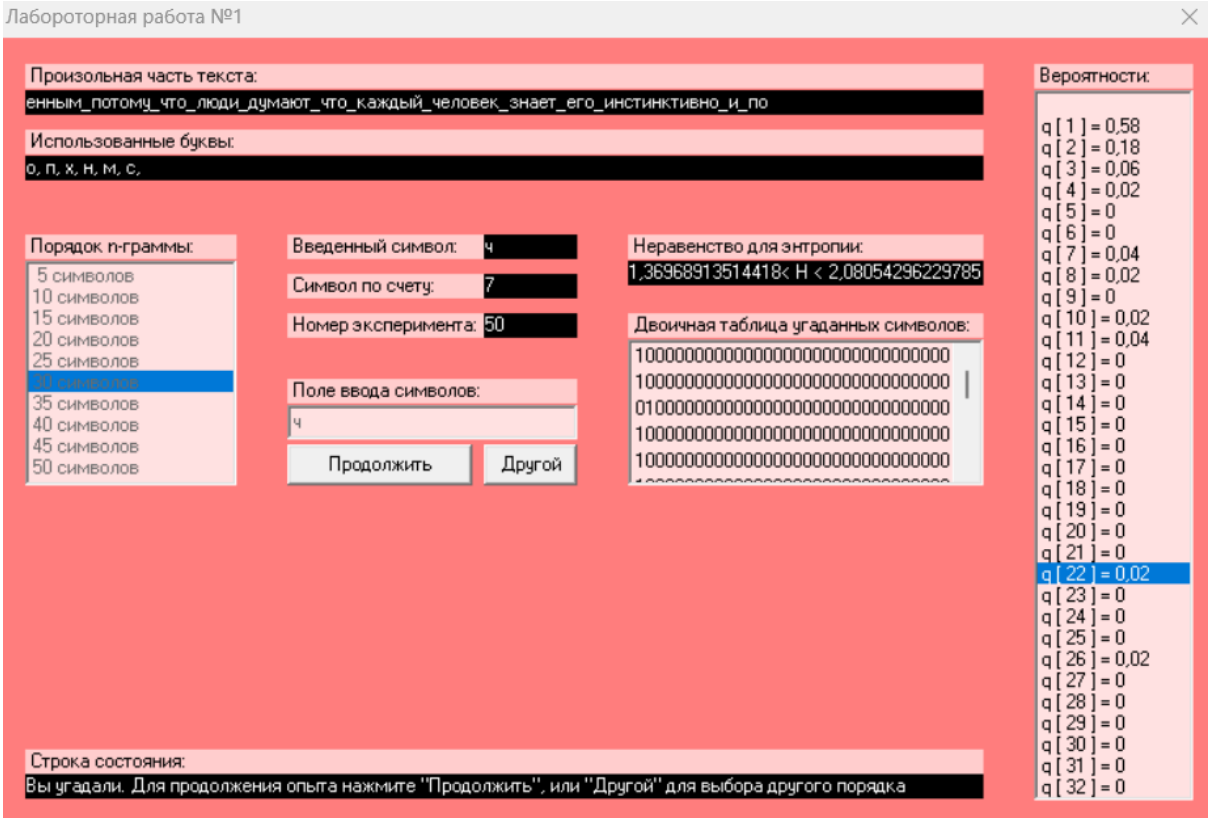
H10


Лабороторная работа №1
Произольная часть текста:
ть_обещание_данное_ему_то_не_успеете_вы_и_слово_вымолвить_как_он_станет_жал
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: и
Символ по счету: 1
Номер эксперимента: 50
Поле ввода символов:
и
Продолжить
Другой
Неравенство для энтропии:
1,87436345865577< H < 2,63433428621155
Двоичная таблица угаданных символов:
Вероятности:
q [1] = 0,52
q [2] = 0,12
q [3] = 0,06
q [4] = 0
q [5] = 0,02
q [6] = 0,04
q [7] = 0,04
q [8] = 0
q [9] = 0,02
q [10] = 0,02
q [11] = 0,02
q [12] = 0
q [13] = 0
q [14] = 0
q [15] = 0,02
q [16] = 0
q [17] = 0,04
q [18] = 0,04
q [19] = 0
q [20] = 0
q [21] = 0
q [22] = 0,02
q [23] = 0
q [24] = 0
q [25] = 0
q [26] = 0
q [27] = 0
q [28] = 0
q [29] = 0
q [30] = 0
q [31] = 0
q [32] = 0,02
Строка состояния:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

H20



H30



Оцінка надлишковості:

Формула:

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

	З пробілами	Без пробілів
H0 (максимальна ентропія)	5.087462	5.044394
R для H1	13.58%	11.38%
R для H2 (перетин)	29.64%	24.41%
R для H2 (без перетину)	29.60%	24.54%

Якщо значення ентропій лежать в таких межах:
H(10) - 1.874 < H < 2.545
H(20) - 1.778 < H < 2.545
H(30) - 1.367 < H < 2.081

Відповідно надлишковість:
H(10) - 49.97% < R < 63.17%
H(20) - 49.97% < R < 65.05%
H(30) - 59.10% < R < 73.13%

Висновок: У ході виконання роботи були написані програми для підрахунку частот появи окремих символів і біграм у тексті російською мовою та обчислені значення ентропії H1 і H2. Отримано, що ентропія тексту без пробілів вища, ніж з пробілами, оскільки відсутність пробілів робить розподіл символів рівномірним і менш передбачуваним. Також встановлено, що H2, H1, бо символи в мові не є незалежними — їх поява залежить від контексту (сусідніх символів). За результатами оцінки ентропій вищих порядків H(10), H(20),(30) визначено, що реальна ентропія російської мови становить приблизно 1.5–2.5 біт/символ, а її надлишковість — близько 60–70%, що підтверджує високу структурованість та закономірність мови.