

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:
ФБ-31 Аль-Фітурі Асія
ФБ-31 Гриб Вероніка

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 а) частот букв і б) частот біграм в тексті, а також підрахунку с) 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.

А) частота букв

1. Частота літер

Частота окремої літери обчислюється як відношення кількості появ цієї літери до загальної кількості літер у тексті

$$P(l) = \frac{f(l)}{n}$$

Дані збереглись а таблиці

Створено файл 'crypto_5tables.xlsx' з 7 аркушами.

Частота літери з пробілами

	A	B	C
1	Літера	Кількість	Частота
2	пробіл	173144	0.16767
3	о	95774	0.09275
4	а	74368	0.07202
5	е	73361	0.07104
6	е	73361	0.07104
7	и	53873	0.05217
8	н	52690	0.05102
9	т	51049	0.04944
10	с	44734	0.04332
11	л	43373	0.042
12	р	43076	0.04171
13	в	41001	0.0397
14	к	31695	0.03069
15	м	27873	0.02699
16	д	25559	0.02475
17	у	24698	0.02392
18	п	21472	0.02079
19	г	18252	0.01768
20	ь	17094	0.01655
21	ь	17094	0.01655
22	я	16841	0.01631
23	ы	14150	0.0137
24	з	13902	0.01346
25	б	13350	0.01293
26	ч	11986	0.01161
27	й	10369	0.01004
28	ж	8371	0.00811
29	ш	8121	0.00786
30	х	7379	0.00715
31	ю	5223	0.00506
32	ц	3753	0.00363
33	э	2827	0.00274
34	щ	2233	0.00216
35	ф	1050	0.00102
вс			

Частота літерів без пробілів

	A	B	C
1	Літера	Кількість	Частота
2	о	95774	0.11143
3	а	74368	0.08653
4	е	73361	0.08535
5	и	53873	0.06268
6	н	52690	0.0613
7	т	51049	0.05939
8	с	44734	0.05205
9	л	43373	0.05046
10	р	43076	0.05012
11	в	41001	0.0477
12	к	31695	0.03688
13	м	27873	0.03243
14	д	25559	0.02974
15	у	24698	0.02874
16	п	21472	0.02498
17	г	18252	0.02124
18	ь	17094	0.01989
19	я	16841	0.01959
20	ы	14150	0.01646
21	з	13902	0.01617
22	б	13350	0.01553
23	ч	11986	0.01395
24	й	10369	0.01206
25	ж	8371	0.00974
26	ш	8121	0.00945
27	х	7379	0.00859
28	ю	5223	0.00608
29	ц	3753	0.00437
30	э	2827	0.00329
31	щ	2233	0.0026
32	ф	1050	0.00122
33			

Біграми які перетинаються з пробілом

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	
а	9E-06	0.00059	0.00329	0.00093	0.00205	0.00106	0.00104	0.0036	9.3E-05	0.00073	0.0042	0.0071	0.00341	0.00386	3.5E-05	0.00061	0.00497	0.00461	0.00499	3.9E-05	0.00017	0.00078	2.3E-05	0.00076	0.00153	0.00003	0	0	0	0	0	0.00076	0.00179	0.0187	
б	0.0009	3E-06	9.5E-05	1.1E-05	7.1E-05	0.00188	0.00188	5E-06	1E-06	0.00085	0	0.00011	0.0008	5.4E-05	0.00026	0.00188	0	0.00126	0.00033	0	0.0012	0	0.00004	4E-06	0	1.1E-05	7.1E-05	0.00011	0.0026	0.00011	0	6E-06	0.00042	0.00023	
в	0.00663	8E-06	4.2E-05	2.5E-05	0.00024	0.00066	0.00066	0	0.00503	0.00267	0	0.00011	0.00066	0.0004E	0.00069	0.00064	0.00018	0	0.0009	0.00023	0.00025	0.00076	0	9.4E-05	1.9E-05	7.2E-05	0.00098	6E-06	0.00024	0.00327	0.00024	0	9E-06	0.00023	0.00561
г	0.00195	2E-06	1.6E-05	4E-06	0.0001	0.00222	0.0022	3E-06	0	0.00085	0	7.3E-05	0.00105	8E-06	0.00028	0.00752	0	0.00089	4.6E-05	0.0004	0.0009	0	0	1E-06	1.9E-05	3E-06	0	0	0	1E-06	0	0.00106	0.00106		
д	0.00465	3.4E-05	0.00143	1.9E-05	4.4E-05	0.00469	0.00469	1.4E-05	1.1E-05	0.00219	0	0.00011	0.00567	0.00014	0.00185	0.00386	9.7E-05	0.00116	0.0003	0.00015	0.00147	0	3.8E-05	0.0002	2.8E-05	4.1E-05	8E-06	0.0004	0.00042	0.00071	0	2.3E-05	0.00026	0.00083	
е	7.5E-05	0.00104	0.00179	0.00039	0.00259	0.00124	0.00124	0.00075	0.00092	8.9E-05	0.00194	0.00142	0.00563	0.00386	0.00747	0.00016	0.00076	0.00684	0.05008	0.00505	0.00014	1.4E-05	0.0005	0.00042	0.00088	0.00064	0.00052	0	0	0	0.00013	0.00016	0.01777		
ж	0.00127	5.2E-05	0	9E-06	0.00088	0.00327	0.00327	0.00001	0	0.00123	0	7.7E-05	2.4E-05	4E-06	0.00061	2.3E-05	0	3E-06	9E-06	0	0.00033	0	0	4.8E-05	0	0	2.9E-05	2.9E-05	0	0	0	0	0.00023		
з	0.00511	0.00015	0.00072	0.0004	0.00078	0.00022	0.00022	0.00005	6.6E-05	0.00024	0	0.00009	0.00018	0.00024	0.00173	0.00095	1.5E-05	0.00032	8.2E-05	8E-06	0.00034	0	0	4E-06	5E-06	1E-06	0	0.00017	0.00031	0.00017	0	0.00031	0.00139		
и	0.00007	0.00052	0.00162	0.00005	0.00044	0.00144	0.0019	0.0019	0.00032	0.00204	0.00048	0.00152	0.0019	0.00502	0.00043	0.00345	0.00017	0.00013	0.00083	0.00198	0.00452	7E-06	2.9E-05	0.00231	0.00086	0.00125	0.00051	0.00016	0	0	2E-06	0.00024	0.00096	0.01487	
я	8E-06	1.3E-05	3E-06	0	0.00022	1.3E-05	1.3E-05	0	2E-06	1E-06	0	0.00004	7E-06	3E-05	0.00038	9.1E-05	2E-06	2E-06	0.00046	0.0006	0	1E-06	0	8.7E-05	0.00017	0.00005	0	0	0	1E-06	0	0.000785			
к	0.00706	0	9.5E-05	3E-06	0	0.00048	0.00048	1.3E-05	2E-06	0.00237	0	7E-06	1E-06	0.00075	0.01048	0	0.00173	0.00005	0.00041	0.0012	0	3E-06	6E-06	0	1.3E-05	0	3E-06	0	0	4E-05	0	0.00439			
л	0.00748	4.1E-05	2.2E-05	0.00007	7.7E-05	0.00394	0.00394	0.00031	6E-06	0.00625	0	0.00025	4.8E-05	1.5E-05	0.00031	0.00069	5.8E-05	0	0.00119	4.3E-05	0.00153	2E-06	3E-06	3E-06	9.5E-05	9E-06	7E-06	0.00045	0.00066	0.00045	0	0.0012	0.00083	0.0074	
м	0.00034	4.7E-05	2E-06	6.1E-05	0	0.00346	0.00346	0	1E-06	0.00225	0	6.4E-05	0.00013	9E-06	0.00153	0.00495	5.6E-05	0.00011	9.3E-05	0.00004	0.00233	4E-06	0	1.8E-05	0.00003	2E-06	4.1E-05	5.2E-05	0.00075	5.2E-05	0.00011	1E-06	0.00048	0.00774	
н	0.01027	6E-06	1.6E-05	5.8E-05	0.00027	0.00083	0.00053	0.00021	2.7E-05	0.00613	0	0.00017	1.7E-05	2E-05	0.00266	0.00793	0	0.00015	0.00112	0.00051	0.00276	2.5E-05	2E-06	0.00047	0.00014	2E-06	0.00017	0.0008	0.00256	0.0008	0	0.00016	0.00169	0.00316	
о	0.1E-06	0.00323	0.00077	0.00487	0.00447	0.00123	0.00213	0.00169	0.00107	0.00089	0.00406	0.00044	0.00759	0.00444	0.00571	0.00015	0.00108	0.00677	0.00593	0.00067	4.6E-05	3.5E-05	0.0006	0.00001	0.00162	0.0013	0.00016	0	0	9.7E-05	0.00035	0.00067	0.01912		
п	0.00139	0	0.0005	0	0.00018	0.00168	0	0	0.00074	0	3.6E-05	0.00063	0	9.6E-05	0.00787	0.00165	0.00084	9E-06	0.00078	0.00068	0	1E-06	6E-06	2E-06	0	0.00013	0.00028	0.00013	0	3E-06	0.00021	5.7E-05			
р	0.00574	8.4E-05	0.00048	0.00009	0.00033	0.00024	0.00055	0.00059	0.00029	4.9E-05	0.00708	0	0.00025	0.00081	0.00019	0.0008	0.00826	0.00065	0.00061	0.00087	0.00063	0.00238	6.8E-05	0.00013	0.00092	6.3E-05	0.00017	3.5E-05	0.00071	0.00114	0.00071	3.4E-05	0.00012	0.00102	0.0006
с	0.00143	4.8E-05	0.00061	2.5E-05	0.0003	0.00035	0.00035	0.00035	3.9E-05	1.1E-05	0.00178	0	0.00483	0.00281	0.00007	0.00072	0.00272	0.00194	0.00016	0.00055	0.00048	0.00071	9E-06	0.00016	4.1E-05	0.00029	4.6E-05	0	0.00296	0.00031	0.00298	0	0.00014	0.00296	0.00342
т	0.00551	1.8E-05	0.00331	9E-06	0.00012	0.0061	0.0061	0	3E-06	0.00371	0	0.00048	0.00019	1.1E-05	0.00106	0.0115	3E-05	0.00236	0.00097	0.00019	0.00143	6E-06	1.2E-05	8.8E-05	9.7E-05	1.6E-05	2.2E-05	0.00538	0.00019	0.00538	0	7.3E-05	0.00038	0.00514	
у	0.0003	0.0006	0.001	0.00218	0.00179	0.00021	0.00021	0.00121	0.00045	0.00037	0.00019	0.00078	0.00176	0.00091	0.00015	4E-06	0.00059	0.00031	0.001	0.00122	3E-06	1.4E-05	0.00038	5E-06	0.00077	0.00062	0.00017	0	0	3.2E-05	0.00096	9.2E-05	0.00076		
ф	0.00011	0	0	1E-06	0	5.7E-05	5.7E-05	0	0	0.00011	0	7.8E-05	0	2E-06	0.00011	0	0.00035	4E-06	3E-06	2.8E-05	3E-06	0	0	0	0	0	0	0	4E-06	1.4E-05	4E-06	0	3.1E-05	0	0.00011
х	0.0006	0	0.00015	1E-06	1E-06	3.8E-05	3.8E-05	0	0	0.00013	0	0	5.2E-05	2.7E-05	0.00014	0.00207	0	0.00021	2.5E-05	0.00002	0.00012	0	0	0	4E-06	2E-06	0	1E-06	0	1E-06	2E-06	0	0	0.00356	
ц	0.00072	0	4.6E-05	0	0	0.00075	0.00075	0	0	0.00017	0	1.2E-05	3E-06	0	0	0.0103	0	0	0	0	0	7E-06	0.00027	0	0	1E-06	0	0	0	0	0.0002	0	0	0.00042	
ч	0.00178	0	2E-06	0	0	0.00033	0.00033	0	0.00124	0	0.00016	2.5E-05	0	0.00064	4.7E-05	0	1.1E-05	0	0.00399	0.00054	0	0	0	0	5E-06	0.00014	0	0.00028	0	0.00028	0	0	0	0.0001	
ш	0.00105	0	3.6E-05	0	0	0.00277	0.00277	0	0	0.00152	0	0.00029	0.00047	3E-06	0.00033	0.00032	0.0002	3E-06	0	1.5E-05	0.0003	2E-06	0	0	0	1E-06	0	0.0004	0	0.0004	0	0	0	0.00017	
щ	0.00035	0	0	0	0	0.00092	0.00092	0	0	0.00064	0	0	0	0	3.9E-05	1.1E-05	0	3E-06	0	0	0	8.3E-05	0	0	0	0	0	0	0	4.6E-05	0	4.6E-05	0	1E-06	7.7E-05
ъ	1E-06	0.00011	1.5E-05	1.9E-05	0.0007	0.0007	0.0007	0	0.00014	7.6E-05	0	0.00092	0	0.00021	0.00021	0.00024	2.4E-05	4E-06	6E-06	0.00077	0.00013	0	2E-06	0	5.9E-05	6.4E-05	0.00042	7.7E-05	0	0	0	0.00048	0.00035	0.01102	
ы	0	0.00021	0.00068	5.2E-05	7.6E-05	0.00062	0.00062	3.3E-05	9.6E-05	1.2E-05	0.00153	8.7E-05	0.00125	0.00058	0.00027	0	9.8E-05	0.00022	0.0006	0.00062	1E-06	0	0.0008	7E-06	6.00E-05	0.00045	4E-06	0	0	0	1E-06	0.00043	0.00094		
ь	1E-06	0.00011	1.5E-05	1.9E-05	0.0007	0.0007	0.0007	0	0.00014	7.6E-05	0	0.00092	0	0.00021	0.00021	0.00024	2.4E-05	4E-06	6E-06	0.00077	0.00013	0	2E-06	0	5.9E-05	6.4E-05	0.00042	7.7E-05	0	0	0	0.00048	0.00035	0.01102	
э	0	0	1E-06	0.00001	5E-06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2E-06	0	1E-06	6.7E-05
ю	9E-06	0.00051	1.3E-05	2E-06	0.00026	1E-06	1E-06	0	0	0.00012	0	0	0	0	0	0	0	0.00019	0.00002	0.00029	0.0003	0	1E-06	0	0	0	0.00012	0.00011	0.00021	0	0	0	4.6E-05	0	0.0001
я	0	8E-06	0.00044	8.6E-05	0.00042	9.4E-05	9.4E-05	0.00012	0.00015	7.2E-05	3.6E-05	0.00015	0.00067	0.00027	0.00066	0	4.5E-05	7.3E-05	0.00049	0.00122	0	0	0.00014	2.4E-05	0.00011	0.00001	0.00015	0.00015	0	0	0	9.5E-05	8.8E-05	0.01071	
пробілі	0.00273	0.00561	0.01683	0.00053	0.00746	0.04049	0.04049	0.00199	0.00425	0.00793	1E-06	0.01157	0.00363	0.00078	0.001356	0.01024	0.01476	0.00353	0.01478	0.00663	0.00433	0.0.													

Біграми з пробілом без перетину

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	
2	a	0.00001	0.00567	0.00342	0.00097	0.00265	0.00108	0.00108	0.00102	0.00385	0.000093	0.00707	0.00114	0.007215	0.03465	0.03885	0.00039	0.00587	0.00483	0.004631	0.005024	0.000249	0.000141	0.000802	0.00027	0.00765	0.000302	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
3	a	0.00055	0.00004	0.00004	0.00000	0.00001	0.001892	0.01892	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
4	b	0.06697	0.00006	0.00003	0.00001	0.00001	0.00598	0.00598	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
5	c	0.0184	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
6	d	0.04586	0.0004	0.001427	0.000015	0.00005	0.00497	0.00497	0.00014	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
7	e	0.00072	0.01083	0.01697	0.00295	0.02601	0.00122	0.00122	0.00705	0.00889	0.000077	0.01952	0.01396	0.05661	0.03837	0.07592	0.00169	0.00668	0.06876	0.05109	0.04958	0.00138	0.00014	0.004486	0.00455	0.00893	0.00631	0.00536	0.0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
8	e	0.00077	0.01083	0.01697	0.00295	0.02601	0.00122	0.00122	0.00705	0.00889	0.000077	0.01952	0.01396	0.05661	0.03837	0.07592	0.00169	0.00668	0.06876	0.05109	0.04958	0.00138	0.00014	0.004486	0.00455	0.00893	0.00631	0.00536	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
9	ж	0.01323	0.00054	0.00000	0.00000	0.00000	0.003875	0.00385	0.00285	0.00008	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
10	z	0.0526	0.00138	0.00067	0.00395	0.00084	0.00217	0.00217	0.00041	0.00004	0.00074	0.00217	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
11	z	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
12	я	0.000012	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
13	к	0.07978	0.00000	0.00000	0.00000	0.00000	0.00048	0.00048	0.00015	0.00000	0.00002	0.002431	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
14	л	0.07307	0.00037	0.00017	0.00066	0.00079	0.00399	0.00399	0.00289	0.00008	0.00693	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
15	м	0.07874	0.00054	0.00002	0.00066	0.00000	0.00344	0.00344	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
16	н	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
17	о	0.000002	0.00153	0.005743	0.04956	0.04538	0.00124	0.00124	0.01687	0.00114	0.00093	0.004011	0.002342	0.00767	0.00346	0.00505	0.00159	0.00161	0.00653	0.05514	0.0673	0.00005	0.00029	0.00628	0.00093	0.01693	0.00238	0.00151	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
18	п	0.01367	0.00000	0.00000	0.00000	0.00000	0.001972	0.00172	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
19	р	0.0685	0.00103	0.00509	0.00918	0.0028	0.03567	0.03567	0.00267	0.00041	0.00732	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
20	c	0.01408	0.00054	0.00058	0.00021	0.00304	0.00345	0.00345	0.00004	0.00000	0.00784	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
21	t	0.05499	0.00019	0.00348	0.00006	0.00112	0.00616	0.00616	0.00000	0.00000	0.00375	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
22	ф	0.000587	0.00000	0.001284	0.00805	0.00303	0.01197	0.01197	0.00000	0.00000	0.00169	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
23	0.00126	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
24	x	0.00567	0.00014	0.000002	0.00000	0.00000	0.00005	0.00005	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
25	ц	0.00742	0.00004	0.00000	0.00000	0.00000	0.000759	0.000759	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
26	ч	0.01724	0.00000	0.00000	0.00000	0.00000	0.003599	0.003599	0.00000	0.00000	0.001269	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
27	ш	0.00195	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
28	щ	0.00329	0.00000	0.00000	0.00000	0.00000	0.000883	0.000883	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
29	b	0.00103	0.00105	0.00021	0.00001	0.00077	0.00177	0.00177	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
30	y	0.00198	0.00688	0.00054	0.00007	0.00068	0.00628	0.00043	0.00041	0.00001	0.00001	0.001569	0.00091	0.001313	0.00097	0.00277	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
31	b	0.00103	0.00015	0.00001	0.00001	0.00077	0.0017																													

Без пробілу без перетину

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
1		а	б	в	г	д	е	є	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ь	ы	ь	э	ю	я
2		0.000428	0.001364	0.005997	0.001983	0.00333	0.002131	0.002131	0.001459	0.004814	0.001156	0.00088	0.006611	0.009019	0.006392	0.006529	0.001401	0.002571	0.006346	0.007397	0.006797	0.000645	0.000279	0.001112	0.000072	0.001478	0.001964	0.000358	0	0	0	0.000435	0.001119	0.002464
3		0.001077	0.000002	0.000135	0	0.00003	0.002276	0.002276	0.000007	0.000002	0.00107	0	0.000133	0.00097	0.000079	0.0000328	0.002376	0.000014	0.001594	0.000156	0.000007	0.001426	0.000002	0.000061	0.000005	0	0.000014	0.000086	0.000128	0.00302	0.000128	0.000079	0.000007	0.00057
4		0.0081	0.00172	0.000389	0.000344	0.000068	0.00733	0.00733	0.000035	0.000814	0.001449	0	0.000763	0.001056	0.000365	0.001273	0.008447	0.000914	0.001317	0.005255	0.000842	0.001008	0.000009	0.000137	0.000028	0.000289	0.001233	0.000007	0.000289	0.003828	0.000289	0.000305	0.000019	0.000321
5		0.002669	0.000049	0.000126	0.000094	0.001287	0.002716	0.002716	0.000016	0.000037	0.001035	0	0.000144	0.001226	0.000044	0.00037	0.00914	0.000109	0.000845	0.000168	0.000009	0.001047	0.000005	0.000002	0.000005	0.000033	0.000002	0.000002	0	0	0	0	0.00012	
6		0.005682	0.000074	0.001878	0.000074	0.000074	0.00554	0.00554	0.000016	0.000079	0.002646	0	0.000219	0.000717	0.000128	0.002355	0.004042	0.000207	0.001357	0.000449	0.00024	0.001757	0.000005	0.000004	0.000249	0.000091	0.000056	0.000012	0.000831	0.000484	0.000831	0.000019	0.000026	0.000323
7		0.000335	0.002024	0.004682	0.004251	0.003916	0.001852	0.001852	0.001152	0.001768	0.001017	0.002183	0.002855	0.007283	0.00622	0.010532	0.00155	0.003044	0.00884	0.00833	0.006916	0.0008	0.000081	0.000787	0.00054	0.001617	0.000896	0.000589	0	0	0	0.000261	0.000165	0.000437
8		0.000335	0.002024	0.004682	0.004251	0.003916	0.001852	0.001852	0.001152	0.001768	0.001017	0.002183	0.002855	0.007283	0.00622	0.010532	0.00155	0.003044	0.00884	0.00833	0.006916	0.0008	0.000081	0.000787	0.00054	0.001617	0.000896	0.000589	0	0	0	0.000261	0.000165	0.000437
9		0.001501	0.000081	0.000033	0.000033	0.001145	0.003933	0.003933	0.000019	0.000007	0.001443	0	0.000102	0.000033	0.000019	0.000807	0.000058	0.000023	0.000007	0.000021	0.000014	0.000396	0	0	0.000051	0	0	0.000037	0.000005	0.000005	0	0.000014	0.000014	
10		0.00619	0.000207	0.001022	0.000538	0.001019	0.000258	0.000258	0.000077	0.000105	0.000365	0	0.000285	0.001247	0.000396	0.002169	0.000789	0.00023	0.000412	0.000226	0.000095	0.000496	0.000009	0	0.000007	0.000004	0.000019	0	0.000209	0.000363	0.000209	0.000337	0	0.000361
11		0.000256	0.001359	0.004284	0.001136	0.002388	0.002841	0.002841	0.00057	0.002286	0.001415	0.001841	0.00333	0.006478	0.003567	0.000549	0.001503	0.001971	0.001419	0.004086	0.006192	0.000442	0.000095	0.000293	0.001031	0.00199	0.00071	0.000179	0	0	0.00024	0.000286	0.001485	
12		0.000158	0.000037	0.000856	0.000291	0.000798	0.000168	0.000168	0.000182	0.00023	0.000477	0	0.000877	0.00023	0.000517	0.000546	0.000672	0.000977	0.000309	0.001515	0.001073	0.000219	0.000077	0.000074	0.000163	0.000484	0.000163	0.000007	0	0	0	0.000091	0.000014	0.0001
13		0.009603	0.000291	0.000686	0.000182	0.000258	0.000691	0.000691	0.000133	0.000109	0.003106	0	0.000454	0.001087	0.000237	0.001282	0.01297	0.000375	0.002152	0.000586	0.000775	0.001559	0.000028	0.000056	0.000012	0.000337	0.000056	0.000002	0.000002	0	0.000002	0.000114	0.000049	0.000109
14		0.009126	0.000333	0.00081	0.0008	0.000363	0.005117	0.005117	0.0004	0.000226	0.0078	0	0.001485	0.000461	0.000333	0.01042	0.006132	0.000216	0.002064	0.00034	0.001978	0.000049	0.000063	0.000012	0.000377	0.000067	0.000012	0.005373	0.000752	0.005373	0.00004	0.001519	0.002234	
15		0.03849	0.000335	0.000912	0.000389	0.000517	0.004468	0.004468	0.000121	0.000138	0.003216	0	0.000828	0.000351	0.00285	0.006132	0.001083	0.001298	0.001022	0.000377	0.00393	0.000054	0.000065	0.000033	0.000361	0.000037	0.000061	0.000054	0.000949	0.000054	0.000216	0.000002	0.000742	
16		0.012472	0.000235	0.004244	0.000147	0.000353	0.010152	0.010152	0.00033	0.000184	0.007602	0	0.000421	0.000088	0.0002	0.003507	0.009715	0.000465	0.001475	0.001661	0.000777	0.003451	0.000026	0.00003	0.000358	0.00003	0.000198	0.000999	0.00018	0.000998	0.000042	0.000182	0.002145	
17		0	0.00023	0.00471	0.009641	0.006455	0.006504	0.003207	0.003207	0.00245	0.001983	0.001962	0.004905	0.004198	0.009582	0.006485	0.001882	0.003383	0.008472	0.008752	0.008989	0.000668	0.000116	0.000898	0.000163	0.002853	0.001664	0.000216	0	0	0	0.000496	0.000458	0.001368
18		0.001587	0	0.000007	0.000005	0.000005	0.002003	0.002003	0.000002	0.000002	0.000919	0	0.000037	0.000712	0.000005	0.00112	0.009278	0.00004	0.008275	0.00003	0.000154	0.000803	0	0	0	0.000007	0	0	0.000154	0.00033	0.000154	0.000002	0	0.000249
19		0.008133	0.000128	0.000738	0.00114	0.000356	0.006543	0.006543	0.000354	0.000079	0.008624	0	0.000424	0.00101	0.000289	0.010129	0.010071	0.000184	0.00084	0.001101	0.000824	0.00299	0.000084	0.000142	0.001096	0.000121	0.000221	0.000035	0.000852	0.001378	0.000852	0.000007	0.000128	0.00128
20		0.001864	0.0002	0.002292	0.000219	0.000338	0.003809	0.003809	0.000144	0.000079	0.002243	0	0.006176	0.00357	0.001096	0.001382	0.003509	0.002783	0.000261	0.000856	0.011476	0.00101	0.000037	0.000195	0.000063	0.000389	0.000093	0	0.000686	0.000386	0.000686	0.000091	0.00017	0.000656
21		0.006602	0.000298	0.004856	0.000188	0.004444	0.007593	0.007593	0.000074	0.000128	0.004712	0	0.000912	0.00377	0.000349	0.001924	0.01428	0.000554	0.003016	0.00175	0.000493	0.001801	0.000023	0.000049	0.000126	0.000316	0.000337	0.000023	0.006427	0.001475	0.006427	0.00142	0.000093	0.000507
22		0.000577	0.000926	0.002087	0.00181	0.002464	0.00043	0.00043	0.00154	0.00077	0.001066	0.00024	0.001631	0.002269	0.001589	0.00093	0.00043	0.001347	0.000538	0.001901	0.001773	0.000163	0.000051	0.000051	0.000009	0.00134	0.000805	0.00023	0	0	0.000116	0.001098	0.000235	
23		0.000147	0.000002	0.000007	0.000002	0.000037	0.00007	0.00007	0.000002	0	0.000156	0	0.000016	0.000116	0.000007	0.000005	0.000135	0.000009	0.000375	0.000009	0.000007	0.00004	0.000012	0	0	0	0.000007	0.000012	0.000007	0.000002	0.000033	0.000005		
24		0.000789	0.000151	0.000577	0.000116	0.000312	0.000119	0.000119	0.00004	0.000088	0.000379	0	0.000263	0.000175	0.000202	0.000061	0.002736	0.000433	0.000361	0.000442	0.000135	0.000265	0.000035	0.000009	0.000007	0.000095	0.000019	0.000007	0.000002	0	0.000002	0.000042	0.000002	0.000061
25		0.000873	0.000014	0.000116	0.000012	0.000016	0.000908	0.000908	0.000002	0.000012	0.000254	0	0.000007	0.000005	0.000021	0.000033	0.001287	0.000063	0.000009	0.000054	0.000023	0.000326	0.000009	0.000005	0.000002	0.000012	0	0	0.000235	0	0.000009	0	0.000002	
26		0.002099	0.000005	0.000019	0	0.000009	0.004223	0.004223	0	0	0.001487	0	0.000226	0.000042	0	0.000798	0.000081	0.000016	0.000019	0.000012	0.003725	0.000721	0	0	0	0.000012	0.000151	0	0.00034	0.000012	0	0	0	
27		0.001124	0.000028	0.000037	0.000012	0.000026	0.003379	0.003379	0.000002	0.000007	0.001829	0	0.000377	0.000561	0.000019	0.000377	0.000421	0.000282	0.000009	0.000016	0.000012	0.000356	0.000002	0	0	0.000005	0.000002	0	0.000507	0.000007	0.000002	0	0	
28		0.000414	0.000002	0.000005	0	0.000002	0.00115	0.00115	0	0.000002	0.000821	0	0.000012	0	0.000005	0.000051	0.000028	0.000009	0.000005	0.000007	0.000002	0.000093	0	0	0	0.000002	0	0	0.000051	0	0	0	0	
29		0.0002	0.000461	0.001613	0.000349	0.0006	0.00134	0.00134	0.000114	0.000572	0.000935	0	0.002036	0.000242	0.000775	0.002355	0.000882	0.001163	0.000216	0.002192	0.000663	0.000347	0.000047	0.000102	0.000086	0.000524	0.000551	0.000088	0	0	0.000179	0.000603	0.000593	
30		0.000079	0.000463	0.001412	0.000226	0.000351	0.000924	0.000924	0.000098	0.000305	0.000356	0.001878	0.000356	0.001657	0.001478	0.000933	0.000461	0.000719	0.000347	0.001287	0.000994	0.000233	0.000028	0.001075	0.000021	0.000268	0.000582	0.000005	0	0	0.000054	0	0.000109	
31		0.0002	0.000461	0.001613	0.000349	0.0006	0.00134	0.00134</																										

Надлишковість джерела відкритого тексту (мови) дорівнює $R = 1 - \frac{H_{\infty}}{H_0}$ і

характеризує величину можливого ущільнення тексту деякою схемою кодування символів без втрати його змісту.

H_{∞} — це ентропія «нескінченного порядку» (умовна ентропія, яку оцінюють через експеримент з n-грамами, наприклад $H(10), H(20), H(30)$)

$H_0 = \log_2 m$, де m — кількість символів в алфавіті (для російської мови без пробілу ≈ 32 , з пробілом ≈ 33).

$H(10)$

Ентропія для H_{10} : $2.028 < H < 2.894$

Надлишковість (R) $0.426 < R < 0.598$

Лабораторная работа №1

Произвольная часть текста:
онцов_я_сам_такой_же_то_есть_мне_самому_не_удается_как_следует_соблюдать_ес

Использованные буквы:
к, т, о, г, л, ы, и, _, ч, н,

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

Введенный символ: а

Символ по счету: 11

Номер эксперимента: 4

Неравенство для энтропии:
 $2.68365034650669 < H < 2$

Двоичная таблица угаданных символов:
01000000000000000000000000000000
00000100000000000000000000000000
10000000000000000000000000000000
00000000001000000000000000000000

Вероятности:
q[1] = 0,25
q[2] = 0,25
q[3] = 0
q[4] = 0
q[5] = 0
q[6] = 0,25
q[7] = 0
q[8] = 0
q[9] = 0
q[10] = 0
q[11] = 0,25
q[12] = 0
q[13] = 0
q[14] = 0
q[15] = 0
q[16] = 0
q[17] = 0
q[18] = 0
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0
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

Поле ввода символов:
а

Продолжить Другой

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

тут ми багато разів вгадували що може бути після с...

Надлишковість (R) $0.538 < R < 0.712$

[illegible]

H(30)

Ентропія $1.484 < H < 2.159$

Надлишковість (R) $0.572 < R < 0.706$

Порядок n-грам	Ентропія (H)	Надлишковість (R)
10 символів	$2.028 < H < 2.894$	$0.426 < R < 0.598$
20 символів	$1.453 < H < 2.331$	$0.538 < R < 0.712$
30 символів	$1.484 < H < 2.159$	$0.572 < R < 0.706$

Висновок:

У роботі на основі книги російською мовою підраховано частоти літер і біграм для двох варіантів (з пробілом/без пробілу) та побудовано матриці біграм з перетином і без, що загалом дало 7 таблиць результатів.

За програмою CoolPinkProgram було досліджено ентропію та надлишковість відкритого тексту за допомогою n-грамного аналізу. Метою було визначити міру неупорядкованості тексту (ентропію H) та частку повторюваної інформації (надлишковість R).

Отримані результати:

$$H10: 2.028 < H < 2.894$$

$$R10: 0.426 < R < 0.598$$

$$H20: 1.452 < H < 2.330$$

$$R20: 0.538 < R < 0.712$$

$$H30: 1.483 < H < 2.159$$

$$R30: 0.572 < R < 0.706$$

Аналіз показав, що зі збільшенням порядку n-грам ентропія зменшується, а надлишковість зростає.