



Міністерство освіти і науки України НТУУ «Київський
політехнічний інститут» Фізико-технічний інститут

КРИПТОГРАФІЯ

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела
відкритого тексту.

Перевірив:
Чорний О. М.

Виконали:
студенти 3 курсу ФТІ
групи ФБ-72
Башкиров І. О.
Давидюк П. П.

Київ – 2019

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. Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.
2. 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.
3. За допомогою програми CoolPinkProgram оцінити значення $(10) H$, $(20) H$, $(30) H$.
4. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Результати:

Монограми:

З пробілом

‘:0.15744762916879637
о:0.0933214293902421
а:0.07242164773409253
е:0.06865626150553412
и:0.05818546684990108
н:0.05420564780961603
т:0.05085698949818676
л:0.04358785204405718
с:0.04282317654563575
р:0.040193340175403346
в:0.03970108875225727
к:0.030495312822407823
м:0.02526126412862732
у:0.024944335130163413
п:0.02403265852181618
д:0.023575471583606546
я:0.016104039024074465
г:0.015905789820780024
ь:0.015177527441331048
з:0.015010296820864986
ы:0.014670441043788798
ч:0.013274604816511597
б:0.012995437571056156
й:0.009900322458398012
ж:0.0075267265550087595
ш:0.007459294853207929
х:0.006984575672530078
ю:0.00449364860800738
щ:0.0031719872527110914
ц:0.0028671959605713355
э:0.0026042123235480944
ф:0.0021443281172664266

$H_1 = 4.379945305657554$

Без пробіла

о:0.1107603902392176
а:0.08595506966842471
е:0.08148604630689321
и:0.06905857589896677
н:0.06433504869986954
т:0.06036062713587143
л:0.05173310710769994
с:0.05082553681902216
р:0.04770426333944249
в:0.047120024970187836
к:0.03619396713859255
м:0.02998183258769578
у:0.02960567911707977
п:0.028523637644158818
д:0.02798101625463189
я:0.019113398266492727
г:0.018878102265724415
ь:0.018013749609840816
з:0.017815268629600877
ы:0.017411904056855197
ч:0.015755228133078295
б:0.015423892948322917
й:0.011750394160817614
ж:0.008933244763863656
ш:0.008853212110541102
х:0.00828978223115031
ю:0.005333376017415106
щ:0.0037647360122930154
ц:0.0034029884192750643
э:0.0030908610713170975
ф:0.002545038375657268

$H_1 = 4.452870917974479$

Біграми:

Перетинаються, з пробілом - $H_2 = 3.991013097121348$

Не перетинаються, з пробілом - $H_2 = 3.990569680311937$

Перетинаються, без пробіла - $H_2 = 4,148671405718143$

Не перетинаються, без пробіла - $H_2 = 4.147853620317564$

Перетинаються, з пробілом

$$H_2 = 3.991013097121348$$

	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	ы	ь	э	ю	я
а	0.000013	0.000564	0.002479	0.000887	0.002193	0.001130	0.001086	0.003815	0.000289	0.000623	0.004212	0.008459	0.003060	0.005431	0.000832	0.000870	0.004262	0.003964	0.004920	0.000086	0.000214	0.001121	0.000116	0.000939	0.000867	0.000382	0.000000	0.000000	0.000013	0.000738	0.002080
б	0.000928	0.000012	0.000020	0.000018	0.000007	0.002274	0.000011	0.000001	0.000646	0.000000	0.000154	0.000570	0.000047	0.000317	0.001931	0.000000	0.001064	0.000081	0.000004	0.001222	0.000000	0.000036	0.000001	0.000005	0.000007	0.000237	0.002696	0.000171	0.000000	0.000016	0.000316
в	0.006549	0.000011	0.000111	0.000019	0.000262	0.005531	0.000000	0.000407	0.003471	0.000000	0.000263	0.000619	0.000121	0.000901	0.006266	0.000212	0.000635	0.002084	0.000290	0.000639	0.000000	0.000038	0.000038	0.000104	0.000924	0.000028	0.002611	0.000467	0.000000	0.000000	0.000270
г	0.002081	0.000001	0.000003	0.000000	0.000881	0.000477	0.000000	0.000001	0.000645	0.000000	0.000076	0.001497	0.000013	0.000378	0.007525	0.000001	0.001268	0.000008	0.000020	0.000477	0.000000	0.000000	0.000000	0.000047	0.000001	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
д	0.004209	0.000067	0.000898	0.000013	0.000016	0.003854	0.000058	0.000009	0.002005	0.000000	0.000293	0.000454	0.000125	0.001923	0.003223	0.000100	0.000976	0.000367	0.000192	0.001439	0.000000	0.000034	0.000204	0.000062	0.000154	0.000000	0.000481	0.000565	0.000000	0.000009	0.000409
е	0.000128	0.000839	0.002047	0.002886	0.002840	0.001274	0.000757	0.001386	0.000192	0.001834	0.001455	0.005834	0.003679	0.006422	0.000208	0.001158	0.006538	0.004096	0.005273	0.000156	0.000117	0.000411	0.000340	0.001055	0.000645	0.000572	0.000000	0.000000	0.000000	0.000179	0.000169
ж	0.001276	0.000024	0.000000	0.000013	0.000715	0.002964	0.000009	0.000000	0.001113	0.000000	0.000139	0.000039	0.000008	0.000662	0.000051	0.000000	0.000004	0.000020	0.000003	0.000181	0.000000	0.000000	0.000000	0.000034	0.000000	0.000000	0.000018	0.000000	0.000000	0.000000	0.000000
з	0.005459	0.000165	0.001084	0.000266	0.000848	0.000615	0.000085	0.000086	0.000492	0.000000	0.000111	0.000355	0.000255	0.001407	0.000720	0.000038	0.000256	0.000013	0.000004	0.000452	0.000000	0.000000	0.000000	0.000013	0.000003	0.000000	0.000394	0.000146	0.000000	0.000001	0.000193
и	0.000244	0.000569	0.002715	0.000494	0.001475	0.001829	0.000244	0.002165	0.000552	0.001621	0.002641	0.005979	0.002596	0.003208	0.000409	0.000321	0.001096	0.003001	0.004125	0.000117	0.000345	0.001457	0.000959	0.001698	0.000487	0.000206	0.000000	0.000000	0.000003	0.000272	0.001150
й	0.000000	0.000013	0.000000	0.000011	0.000112	0.000001	0.000000	0.000000	0.000000	0.000000	0.000128	0.000030	0.000038	0.000401	0.000011	0.000001	0.000000	0.000429	0.000396	0.000000	0.000003	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000013
к	0.006951	0.000001	0.000560	0.000004	0.000000	0.000854	0.000057	0.000015	0.002887	0.000000	0.000024	0.000809	0.000003	0.000672	0.000362	0.000001	0.002103	0.000160	0.000712	0.002117	0.000000	0.000003	0.000032	0.000003	0.000011	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
л	0.007242	0.000076	0.000023	0.000109	0.000066	0.004415	0.000210	0.000028	0.006215	0.000000	0.000387	0.000413	0.000051	0.000460	0.006126	0.000055	0.000000	0.001895	0.000189	0.001303	0.000020	0.000007	0.000000	0.000131	0.000015	0.000001	0.000829	0.003605	0.000001	0.000699	0.001380
м	0.003550	0.000015	0.000066	0.000074	0.000007	0.003200	0.000000	0.000000	0.002755	0.000000	0.000109	0.000181	0.000085	0.001243	0.002968	0.000182	0.000090	0.000264	0.000073	0.001703	0.000004	0.000000	0.000008	0.000026	0.000011	0.000003	0.000730	0.000045	0.000003	0.000000	0.000517
н	0.009115	0.000005	0.000028	0.000097	0.000836	0.008695	0.000011	0.000055	0.007318	0.000000	0.000255	0.000000	0.000000	0.003187	0.009772	0.000000	0.000038	0.000523	0.000824	0.003604	0.000100	0.000000	0.000322	0.000160	0.000004	0.000166	0.003353	0.000901	0.000000	0.000154	0.001405
о	0.000011	0.003121	0.007239	0.004390	0.004511	0.001558	0.001587	0.001521	0.000655	0.003854	0.002521	0.006294	0.005641	0.005287	0.000297	0.001372	0.006086	0.005587	0.006452	0.000069	0.000297	0.000484	0.000128	0.001688	0.000747	0.000225	0.000000	0.000000	0.000139	0.000388	0.000498
п	0.001655	0.000000	0.000000	0.000000	0.000000	0.002169	0.000000	0.000000	0.001253	0.000000	0.000086	0.000871	0.000000	0.000251	0.009222	0.000059	0.006567	0.000032	0.000173	0.000780	0.000000	0.000000	0.000007	0.000020	0.000016	0.000000	0.000329	0.000090	0.000000	0.000028	0.000364
р	0.007602	0.000065	0.000514	0.000980	0.000249	0.005643	0.000240	0.000045	0.005559	0.000000	0.000386	0.000245	0.000233	0.000943	0.008181	0.000061	0.000059	0.000165	0.000939	0.002476	0.000004	0.000202	0.000063	0.000177	0.000430	0.000042	0.001341	0.000607	0.000000	0.000240	0.001003
с	0.001668	0.000067	0.001459	0.000026	0.000245	0.002979	0.000036	0.000018	0.001773	0.000000	0.003869	0.002475	0.000619	0.000851	0.002727	0.001713	0.000183	0.000913	0.009370	0.000740	0.000020	0.000128	0.000077	0.000314	0.000154	0.000003	0.000318	0.003368	0.000001	0.000096	0.003733
т	0.005582	0.000028	0.001985	0.000003	0.000147	0.005749	0.000000	0.000004	0.003465	0.000000	0.000719	0.000206	0.000028	0.001274	0.013105	0.000108	0.002902	0.000833	0.000051	0.001934	0.000065	0.000012	0.000018	0.000344	0.000005	0.000005	0.001399	0.004568	0.000000	0.000069	0.000430
у	0.000107	0.000614	0.000696	0.000939	0.001776	0.000200	0.001164	0.000281	0.000022	0.000140	0.000948	0.002104	0.000921	0.000486	0.000004	0.000616	0.000963	0.001234	0.001610	0.000004	0.000187	0.000798	0.000008	0.000696	0.000669	0.000255	0.000000	0.000000	0.000015	0.001003	0.000092
ф	0.000290	0.000001	0.000000	0.000000	0.000000	0.000486	0.000000	0.000000	0.000370	0.000000	0.000000	0.000086	0.000000	0.000001	0.000285	0.000000	0.000445	0.000023	0.000016	0.000051	0.000013	0.000000	0.000000	0.000000	0.000000	0.000000	0.000030	0.000013	0.000000	0.000000	0.000000
х	0.000552	0.000001	0.000170	0.000055	0.000000	0.000061	0.000000	0.000000	0.000306	0.000000	0.000007	0.000163	0.000055	0.000328	0.001764	0.000000	0.000235	0.000055	0.000020	0.000178	0.000000	0.000003	0.000000	0.000000	0.000023	0.000000	0.000000	0.000007	0.000003	0.000000	0.000000
ц	0.000623	0.000001	0.000062	0.000000	0.000000	0.000844	0.000000	0.000000	0.000359	0.000000	0.000019	0.000005	0.000000	0.000000	0.000251	0.000000	0.000036	0.000000	0.000004	0.000154	0.000000	0.000000	0.000003	0.000000	0.000000	0.000000	0.000160	0.000000	0.000000	0.000000	0.000000
ч	0.002262	0.000000	0.000001	0.000000	0.000001	0.003755	0.000000	0.000000	0.001571	0.000000	0.000475	0.000026	0.000004	0.000774	0.000062	0.000000	0.000065	0.000000	0.002990	0.000461	0.000000	0.000000	0.000000	0.000000	0.000054	0.000000	0.000000	0.000252	0.000000	0.000000	0.000000
ш	0.000987	0.000000	0.000062	0.000000	0.000000	0.002031	0.000000	0.000000	0.002028	0.000000	0.000382	0.000425	0.000007	0.000370	0.000233	0.000054	0.000005	0.000000	0.000096	0.000374	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000326	0.000000	0.000000
щ	0.000436	0.000000	0.000000	0.000000	0.000000	0.001368	0.000000	0.000000	0.001103	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000003	0.000000	0.000000	0.000142	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000030	0.000000	0.000000	0.000000
ы	0.000000	0.000249	0.000825	0.000102	0.000098	0.000966	0.000070	0.000073	0.000003	0.001763	0.000205	0.001858	0.001247	0.000151	0.000000	0.000131	0.000243	0.000653	0.000554	0.000000	0.000000	0.000906	0.000015	0.000158	0.000496	0.000007	0.000000	0.000000	0.000000	0.000000	0.000015
ь	0.000000	0.000045	0.000038	0.000071	0.000104	0.000855	0.000001	0.000213	0.000061	0.000000	0.000886	0.000000	0.000225	0.001280	0.000007	0.000011	0.000001	0.000703	0.000237	0.000000	0.000012	0.000000	0.000116	0.000057	0.000317	0.000009	0.000000	0.000000	0.000004	0.000343	0.000402
э	0.000000	0.000000	0.000001	0.000004	0.000000	0.000000	0.000000	0.000004	0.000000	0.000009	0.000036	0.000027	0.000009	0.000019	0.000000	0.000011	0.000012	0.000013	0.002411	0.000000	0.000009	0.000023	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
ю	0.000001	0.000170	0.000016	0.000008	0.000235	0.00																									

Не перетинаються, з пробілом

$H_2 = 3.990569680311937$

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	ы	ь	э	ю	я	
а	0.000019	0.000545	0.002541	0.000917	0.002155	0.001160	0.001084	0.003836	0.000297	0.000664	0.004143	0.008464	0.003115	0.005440	0.000027	0.000093	0.004208	0.003935	0.005041	0.000081	0.000237	0.001163	0.000119	0.000933	0.000858	0.000399	0.000000	0.000000	0.000016	0.000701	0.002161
б	0.000941	0.000005	0.000027	0.000016	0.000008	0.002258	0.000013	0.000000	0.000601	0.000000	0.000162	0.000558	0.000046	0.000286	0.001885	0.000000	0.001047	0.000070	0.000003	0.001195	0.000000	0.000030	0.000003	0.000011	0.000011	0.000224	0.002592	0.000173	0.000000	0.000019	0.000316
в	0.006441	0.000013	0.000105	0.000024	0.000224	0.005610	0.000000	0.000391	0.003401	0.000000	0.000267	0.000626	0.000100	0.000955	0.006376	0.000208	0.000599	0.002080	0.000289	0.000691	0.000000	0.000032	0.000035	0.000105	0.000912	0.000035	0.002646	0.000472	0.000000	0.000000	0.000267
г	0.002198	0.000003	0.000003	0.000000	0.000836	0.000480	0.000000	0.000003	0.000696	0.000000	0.000086	0.001492	0.000008	0.000348	0.007596	0.000003	0.001330	0.000011	0.000016	0.000469	0.000000	0.000000	0.000000	0.000059	0.000003	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
д	0.004173	0.000076	0.000877	0.000005	0.000013	0.003906	0.000049	0.000013	0.001942	0.000000	0.000299	0.000429	0.000129	0.001942	0.003153	0.000086	0.000971	0.000297	0.000194	0.001478	0.000000	0.000027	0.000213	0.000046	0.000148	0.000000	0.000486	0.000542	0.000000	0.000013	0.000372
е	0.000140	0.000047	0.002015	0.002843	0.002856	0.001305	0.000753	0.001394	0.000200	0.001826	0.001481	0.005877	0.003698	0.006533	0.000229	0.001214	0.006347	0.004227	0.005160	0.000140	0.000129	0.000413	0.000324	0.000974	0.000650	0.000564	0.000000	0.000000	0.000000	0.000181	0.000213
ж	0.001249	0.000019	0.000000	0.000011	0.000761	0.003040	0.000011	0.000000	0.001076	0.000000	0.000132	0.000038	0.000008	0.000717	0.000049	0.000000	0.000003	0.000027	0.000003	0.000183	0.000000	0.000000	0.000000	0.000019	0.000000	0.000000	0.000000	0.000019	0.000000	0.000000	0.000000
з	0.005462	0.000189	0.001119	0.000254	0.000874	0.000575	0.000089	0.000081	0.000496	0.000000	0.000108	0.000372	0.000235	0.001432	0.000753	0.000035	0.000259	0.000013	0.000003	0.000461	0.000000	0.000000	0.000000	0.000016	0.000005	0.000000	0.000429	0.000154	0.000000	0.000000	0.000205
и	0.000221	0.000539	0.002730	0.000496	0.001546	0.001872	0.000245	0.002169	0.000539	0.001710	0.002535	0.006250	0.002595	0.003212	0.000405	0.000310	0.001098	0.002986	0.004121	0.000127	0.000345	0.001451	0.000968	0.001691	0.000440	0.000229	0.000000	0.000000	0.000003	0.000267	0.001138
й	0.000000	0.000013	0.000000	0.000008	0.0000119	0.000000	0.000000	0.000000	0.000000	0.000000	0.000119	0.000032	0.000040	0.000396	0.000013	0.000003	0.000000	0.000421	0.000351	0.000000	0.000000	0.000000	0.000089	0.000132	0.000151	0.000008	0.000000	0.000000	0.000000	0.000000	0.000013
к	0.007064	0.000003	0.000556	0.000003	0.000000	0.000820	0.000057	0.000016	0.002867	0.000000	0.000030	0.000850	0.000000	0.000637	0.008426	0.000000	0.001993	0.000143	0.000696	0.002090	0.000000	0.000000	0.000046	0.000005	0.000016	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
л	0.007059	0.000086	0.000030	0.000124	0.000067	0.004378	0.000189	0.000022	0.006260	0.000000	0.000386	0.000375	0.000049	0.000467	0.006099	0.000059	0.000000	0.001872	0.000200	0.001362	0.000016	0.000008	0.000000	0.000135	0.000016	0.000000	0.000801	0.003631	0.000000	0.000758	0.001470
м	0.003409	0.000016	0.000067	0.000070	0.000005	0.003161	0.000000	0.000000	0.002646	0.000000	0.000111	0.000181	0.000062	0.001273	0.003069	0.000154	0.000092	0.000275	0.000073	0.001734	0.000005	0.000000	0.000011	0.000013	0.000008	0.000003	0.000701	0.000049	0.000005	0.000000	0.000529
н	0.008963	0.000000	0.000030	0.000086	0.000833	0.008607	0.000008	0.000059	0.007288	0.000000	0.000264	0.000000	0.000000	0.003245	0.009813	0.000000	0.000035	0.000488	0.000785	0.003622	0.000097	0.000000	0.000307	0.000173	0.000000	0.000189	0.003150	0.000893	0.000000	0.000183	0.001287
о	0.000013	0.003266	0.007053	0.004391	0.004577	0.001532	0.001551	0.001486	0.000607	0.003868	0.002541	0.006368	0.005578	0.005335	0.000289	0.001322	0.006012	0.005697	0.006344	0.000086	0.000302	0.000442	0.000124	0.001691	0.000763	0.000224	0.000000	0.000000	0.000143	0.000356	0.000486
п	0.001675	0.000000	0.000000	0.000000	0.000000	0.002112	0.000000	0.000000	0.001227	0.000000	0.000009	0.000879	0.000000	0.000262	0.009314	0.000054	0.006714	0.000022	0.000154	0.000815	0.000000	0.000000	0.000005	0.000022	0.000016	0.000000	0.000351	0.000097	0.000000	0.000027	0.000364
р	0.007577	0.000073	0.000537	0.000904	0.000264	0.005699	0.000248	0.000051	0.005764	0.000000	0.000399	0.000251	0.000227	0.000982	0.008213	0.000076	0.000065	0.000189	0.000944	0.002446	0.000003	0.000200	0.000067	0.000151	0.000437	0.000043	0.001405	0.000593	0.000000	0.000224	0.001038
с	0.001602	0.000067	0.001438	0.000022	0.000254	0.002840	0.000038	0.000022	0.001667	0.000000	0.003916	0.002422	0.000642	0.000887	0.002762	0.001699	0.000186	0.000925	0.009484	0.000685	0.000027	0.000140	0.000078	0.000345	0.000140	0.000003	0.000367	0.003296	0.000003	0.000108	0.003725
т	0.005902	0.000027	0.001918	0.000003	0.000135	0.005756	0.000000	0.000003	0.003474	0.000000	0.000696	0.000216	0.000027	0.001292	0.013079	0.000124	0.002813	0.000839	0.000059	0.001929	0.000076	0.000013	0.000016	0.000313	0.000005	0.000008	0.001378	0.004599	0.000000	0.000070	0.000445
у	0.000086	0.000634	0.000720	0.000944	0.001818	0.000232	0.001117	0.000278	0.000019	0.000143	0.000898	0.002107	0.000869	0.000448	0.000008	0.000626	0.000944	0.001179	0.001621	0.000003	0.000200	0.000801	0.000008	0.000701	0.000653	0.000275	0.000000	0.000000	0.000011	0.000949	0.000092
ф	0.000283	0.000003	0.000000	0.000000	0.000000	0.000467	0.000000	0.000000	0.000345	0.000000	0.000000	0.000086	0.000000	0.000000	0.000267	0.000000	0.000426	0.000019	0.000016	0.000065	0.000019	0.000000	0.000000	0.000000	0.000000	0.000000	0.000030	0.000011	0.000000	0.000000	0.000000
х	0.000521	0.000003	0.000200	0.000057	0.000000	0.000057	0.000000	0.000000	0.000337	0.000000	0.000008	0.000170	0.000051	0.000334	0.001718	0.000000	0.000264	0.000043	0.000024	0.000167	0.000000	0.000000	0.000000	0.000000	0.000027	0.000000	0.000000	0.000011	0.000005	0.000000	0.000000
ц	0.000610	0.000000	0.000067	0.000000	0.000000	0.000863	0.000000	0.000000	0.000359	0.000000	0.000016	0.000005	0.000000	0.000000	0.000259	0.000000	0.000038	0.000000	0.000003	0.000148	0.000000	0.000000	0.000005	0.000000	0.000000	0.000000	0.000132	0.000000	0.000000	0.000000	0.000000
ч	0.002360	0.000000	0.000000	0.000000	0.000000	0.003736	0.000000	0.000000	0.001645	0.000000	0.000483	0.000024	0.000003	0.000796	0.000049	0.000000	0.000070	0.000000	0.002986	0.000394	0.000000	0.000000	0.000000	0.000000	0.000054	0.000000	0.000000	0.000240	0.000000	0.000000	0.000000
ш	0.000955	0.000000	0.000057	0.000000	0.000000	0.001999	0.000000	0.000000	0.002023	0.000000	0.000418	0.000442	0.000005	0.000375	0.000237	0.000051	0.000003	0.000000	0.000100	0.000429	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000334	0.000000	0.000000	0.000000
щ	0.000423	0.000000	0.000000	0.000000	0.000000	0.001378	0.000000	0.000000	0.001009	0.000000	0.000000	0.000000	0.000000	0.000043	0.000008	0.000000	0.000003	0.000000	0.000000	0.000156	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000030	0.000000	0.000000	0.000000
ы	0.000000	0.000240	0.000820	0.000121	0.000086	0.000998	0.000076	0.000073	0.000005	0.001826	0.000197	0.001848	0.001362	0.000140	0.000000	0.000129	0.000245	0.000628	0.000515	0.000000	0.000000	0.000949	0.000013	0.000156	0.000496	0.000008	0.000000	0.000000	0.000000	0.000000	0.000013
ь	0.000000	0.000051	0.000035	0.000062	0.000086	0.000839	0.000000	0.000181	0.000059	0.000000	0.000920	0.000000	0.000237	0.001324	0.000005	0.000016	0.000000	0.000650	0.000251	0.000000	0.000011	0.000000	0.000105	0.000038	0.000332	0.000008	0.000000	0.000000	0.000003	0.000356	0.000402
э	0.000000	0.000000	0.000003	0.000003	0.000000	0.000000	0.000000	0.000003	0.000000	0.000019	0.000038	0.000022	0.000008	0.000013	0.000000	0.000003	0.000008	0.000019	0.002282	0.000000	0.000008	0.000030	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
ю	0.000003	0.000159	0.000022	0.000011	0.000213	0.0000																									

Перетинаються, без пробіла

$$H_2 = 4,148671405718143$$

а	б	в	г	д	е	ж	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я			
а	0.000466	0.001410	0.005052	0.001617	0.003329	0.001766	0.001489	0.005133	0.001825	0.000741	0.006299	0.010446	0.004509	0.000491	0.001172	0.003366	0.005535	0.006673	0.006862	0.000666	0.000336	0.001441	0.000174	0.001591	0.001192	0.000480	0.000000	0.000000	0.000346	0.000888	0.002660	
б	0.001104	0.000016	0.000042	0.000024	0.000010	0.002702	0.000014	0.000005	0.000794	0.000000	0.000192	0.000677	0.000064	0.000386	0.002326	0.000021	0.001265	0.000106	0.000013	0.001461	0.000000	0.000043	0.000002	0.000008	0.000011	0.000282	0.003200	0.000203	0.000050	0.000019	0.000386	
в	0.007864	0.000272	0.000629	0.000411	0.000599	0.006696	0.000069	0.000637	0.004395	0.000003	0.001160	0.000968	0.000466	0.001604	0.007915	0.001095	0.001108	0.003288	0.001000	0.000884	0.000038	0.000077	0.000066	0.000338	0.001148	0.000046	0.003099	0.000554	0.000304	0.000008	0.000379	
г	0.002486	0.000030	0.000059	0.000010	0.001069	0.000576	0.000003	0.000445	0.000828	0.000000	0.000114	0.001782	0.000493	0.000970	0.000075	0.001529	0.000072	0.000040	0.000591	0.000002	0.000003	0.000002	0.000061	0.000002	0.000000	0.000000	0.000000	0.000010	0.000000	0.000002		
д	0.005047	0.000122	0.001199	0.000083	0.000062	0.004608	0.000072	0.000082	0.002487	0.000000	0.000469	0.000583	0.000226	0.0002510	0.000385	0.000296	0.001218	0.000597	0.000311	0.001750	0.000003	0.000046	0.000242	0.000101	0.000194	0.000006	0.000571	0.000671	0.000025	0.00014	0.000498	
е	0.000437	0.001940	0.004383	0.003878	0.004056	0.001766	0.001117	0.002340	0.001458	0.002182	0.002595	0.007299	0.005207	0.009303	0.001393	0.003579	0.000255	0.006731	0.007187	0.000751	0.000227	0.000669	0.000421	0.001783	0.000954	0.000693	0.000000	0.000000	0.000314	0.000216	0.000354	
ж	0.001514	0.000046	0.000026	0.000021	0.000861	0.003534	0.000011	0.000010	0.001341	0.000000	0.000189	0.000050	0.000018	0.000831	0.000080	0.000013	0.000010	0.000043	0.000024	0.000222	0.000000	0.000006	0.000000	0.000042	0.000000	0.000000	0.000000	0.000000	0.000021	0.000014	0.000000	0.000006
з	0.006523	0.000251	0.001461	0.000378	0.001092	0.000748	0.000112	0.000149	0.000655	0.000002	0.000320	0.000450	0.000402	0.001866	0.000962	0.000235	0.000359	0.000198	0.000069	0.000563	0.000003	0.000008	0.000006	0.000064	0.000018	0.000002	0.000467	0.000173	0.000034	0.000002	0.000245	
и	0.000547	0.001389	0.005511	0.001140	0.002451	0.002491	0.000437	0.003235	0.001770	0.001924	0.004184	0.007429	0.003712	0.005598	0.001730	0.002643	0.001742	0.005522	0.005796	0.000656	0.000504	0.001841	0.001173	0.002521	0.000736	0.000264	0.000000	0.000000	0.000250	0.000327	0.001537	
й	0.000195	0.000389	0.000808	0.000413	0.000527	0.000075	0.000150	0.000235	0.000698	0.000000	0.000820	0.000198	0.000472	0.001108	0																	

Не перетинаються, без пробіла

H₂ = 4.147853620317564

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	ы	ь	э	ю	я		
а	0.000435	0.001399	0.005013	0.001591	0.003320	0.001668	0.001559	0.005420	0.001834	0.000723	0.006188	0.010600	0.004549	0.000269	0.001143	0.003358	0.005599	0.006579	0.007033	0.000659	0.000349	0.001473	0.000163	0.001604	0.001136	0.000487	0.000000	0.000000	0.000371	0.000951	0.002708	
б	0.001114	0.000013	0.000051	0.000022	0.000006	0.002625	0.000013	0.000003	0.000820	0.000000	0.000163	0.000640	0.000083	0.000400	0.002343	0.000010	0.001242	0.000128	0.000013	0.001444	0.000000	0.000051	0.000003	0.000010	0.000010	0.000282	0.003313	0.000211	0.000048	0.000003	0.000413	
в	0.007936	0.000295	0.000595	0.000391	0.000599	0.006694	0.000070	0.000647	0.004447	0.000003	0.001095	0.000973	0.000467	0.001597	0.007894	0.001146	0.001076	0.003320	0.000944	0.000893	0.000032	0.000074	0.000080	0.000371	0.001120	0.000048	0.003150	0.000602	0.000323	0.000016	0.000403	
г	0.002555	0.000026	0.000061	0.000006	0.001098	0.000563	0.000000	0.000048	0.000781	0.000000	0.000109	0.001838	0.000019	0.000535	0.000983	0.000086	0.001527	0.000083	0.000038	0.000570	0.000000	0.000006	0.000000	0.000045	0.000000	0.000000	0.000000	0.000000	0.000013	0.000000	0.000003	
д	0.004952	0.000102	0.001233	0.000090	0.000074	0.004581	0.000070	0.000099	0.002567	0.000000	0.000487	0.000599	0.000234	0.002519	0.003938	0.000240	0.001265	0.000618	0.000323	0.001758	0.000003	0.000064	0.000240	0.000118	0.000186	0.000003	0.000563	0.000682	0.000029	0.000019	0.000448	
е	0.000416	0.001898	0.004472	0.003944	0.004075	0.001726	0.001069	0.002347	0.001453	0.002186	0.002558	0.007222	0.005096	0.009197	0.001415	0.003595	0.008394	0.006854	0.007206	0.000733	0.000234	0.000669	0.000419	0.001758	0.000970	0.000672	0.000000	0.000000	0.000311	0.000218	0.000432	
ж	0.001562	0.000048	0.000019	0.000016	0.000887	0.003633	0.000010	0.000006	0.001354	0.000000	0.000189	0.000045	0.000026	0.000845	0.000077	0.000013	0.000010	0.000051	0.000010	0.000240	0.000000	0.000006	0.000000	0.000035	0.000000	0.000000	0.000000	0.000029	0.000019	0.000000	0.000013	
з	0.006460	0.000224	0.001460	0.000317	0.000999	0.000685	0.000122	0.000144	0.000688	0.000000	0.000307	0.000487	0.000352	0.001780	0.000964	0.000250	0.000387	0.000198	0.000067	0.000576	0.000003	0.000006	0.000010	0.000058	0.000022	0.000003	0.000503	0.000128	0.000029	0.000003	0.000227	
и	0.000490	0.001409	0.005497	0.001117	0.002420	0.002391	0.000407	0.003205	0.001783	0.001940	0.004171	0.007549	0.003656	0.005503	0.001716	0.002619	0.001706	0.005551	0.005586	0.000618	0.000448	0.001873	0.001114	0.002567	0.000717	0.000272	0.000000	0.000000	0.000224	0.000336	0.001604	
й	0.000182	0.000327	0.000791	0.000371	0.000535	0.000080	0.000138	0.000259	0.000704	0.000000	0.000826	0.000221	0.000515	0.001133	0.000451	0.001063	0.000346	0.001476	0.000871	0.000221	0.000080	0.000080	0.000134	0.000445	0.000307	0.000026	0.000000	0.000000	0.000086	0.000010	0.000118	
к	0.008327	0.000368	0.001284	0.000102	0.000128	0.001165	0.000189	0.000115	0.003800	0.000000	0.000320	0.001828	0.000189	0.001194	0.010443	0.000432	0.002728	0.000631	0.001031	0.002628	0.000010	0.000061	0.000048	0.000227	0.000045	0.000000	0.000000	0.000000	0.000102	0.000003	0.000051	
л	0.008925	0.000471	0.001101	0.000365	0.000311	0.005442	0.000243	0.000218	0.008186	0.000000	0.001063	0.000544	0.000394	0.001434	0.007760	0.001040	0.000240	0.002945	0.000560	0.001655	0.000074	0.000048	0.000010	0.000528	0.000074	0.000010	0.000951	0.004319	0.000067	0.000826	0.001693	
м	0.004341	0.000384	0.000912	0.000464	0.000352	0.003883	0.000061	0.000285	0.004014	0.000000	0.000627	0.000378	0.000320	0.002270	0.004050	0.001152	0.000298	0.001178	0.000355	0.002308	0.000035	0.000093	0.000022	0.000528	0.000070	0.000013	0.000823	0.000054	0.000125	0.000003	0.000720	
н	0.011025	0.000211	0.000352	0.000195	0.001104	0.010491	0.000048	0.000198	0.008944	0.000000	0.000480	0.000042	0.000118	0.004475	0.011675	0.000503	0.000128	0.000925	0.001088	0.004405	0.000150	0.000051	0.000394	0.000285	0.000022	0.000211	0.003874	0.001060	0.000032	0.000182	0.001719	
о	0.000323	0.004741	0.010958	0.005676	0.006118	0.002430	0.002244	0.002596	0.002190	0.004434	0.004152	0.007782	0.007555	0.008788	0.002193	0.004114	0.007840	0.008829	0.008948	0.000810	0.000480	0.000772	0.000154	0.002916	0.001005	0.000291	0.000000	0.000000	0.000567	0.000461	0.000970	
п	0.001882	0.000003	0.000013	0.000003	0.000000	0.002587	0.000000	0.000003	0.001399	0.000000	0.000115	0.001021	0.000000	0.000307	0.011125	0.000064	0.007645	0.000045	0.000221	0.000877	0.000000	0.000000	0.000010	0.000026	0.000029	0.000000	0.000384	0.000125	0.000000	0.000038	0.000471	
р	0.008791	0.000128	0.000743	0.001184	0.000362	0.006908	0.000256	0.000064	0.006822	0.000000	0.000573	0.000323	0.000304	0.001306	0.009921	0.000288	0.000099	0.000327	0.001095	0.002971	0.000006	0.000218	0.000090	0.000259	0.000531	0.000051	0.001578	0.000685	0.000013	0.000291	0.001255	
с	0.002033	0.000250	0.001962	0.000173	0.000384	0.003499	0.000115	0.000125	0.002420	0.000000	0.004805	0.003035	0.000848	0.001434	0.003422	0.002449	0.000301	0.001325	0.011243	0.001024	0.000042	0.000198	0.000112	0.000413	0.000218	0.000003	0.000394	0.003979	0.000054	0.000112	0.004514	
т	0.006768	0.000298	0.003096	0.000179	0.000381	0.007030	0.000160	0.000227	0.004712	0.000000	0.001265	0.000423	0.000269	0.002180	0.016250	0.000756	0.003669	0.001559	0.000330	0.002526	0.000102	0.000058	0.000038	0.000653	0.000045	0.000010	0.001623	0.000170	0.000099	0.000573		
у	0.000333	0.001008	0.001581	0.001268	0.002343	0.000327	0.001457	0.000509	0.000813	0.000157	0.001514	0.002539	0.001405	0.001479	0.000413	0.001533	0.001287	0.002158	0.002215	0.000195	0.000250	0.000980	0.000019	0.001216	0.000794	0.000269	0.000000	0.000000	0.000125	0.001172	0.000173	
ф	0.000327	0.000003	0.000010	0.000000	0.000000	0.000544	0.000003	0.000000	0.000474	0.000000	0.000000	0.000112	0.000000	0.000010	0.000359	0.000003	0.000547	0.000029	0.000022	0.000058	0.000016	0.000000	0.000000	0.000003	0.000000	0.000000	0.000000	0.000038	0.000013	0.000000	0.000003	
х	0.000723	0.000141	0.000480	0.000141	0.000138	0.000102	0.000032	0.000096	0.000647	0.000000	0.000304	0.000327	0.000192	0.000675	0.002225	0.000467	0.000333	0.000387	0.000166	0.000339	0.000022	0.000038	0.000010	0.000086	0.000083	0.000000	0.000000	0.000003	0.000054	0.000000	0.000026	
ц	0.000762	0.000016	0.000125	0.000000	0.000022	0.001040	0.000000	0.000013	0.000474	0.000000	0.000042	0.000013	0.000013	0.000042	0.000285	0.000022	0.000070	0.000058	0.000016	0.000211	0.000000	0.000003	0.000006	0.000010	0.000000	0.000000	0.000166	0.000000	0.000003	0.000000	0.000010	
ч	0.002702	0.000038	0.000074	0.000022	0.000006	0.004475	0.000003	0.000051	0.001898	0.000000	0.000583	0.000058	0.000006	0.000941	0.000018	0.000064	0.000115	0.000061	0.003531	0.000576	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000003	
ш	0.001223	0.000000	0.000093	0.000003	0.000010	0.002427	0.000000	0.000003	0.002523	0.000000	0.000464	0.000493	0.000010	0.000458	0.000301	0.000086	0.000010	0.000006	0.000102	0.000445	0.000000	0.000006	0.000000	0.000003	0.000000	0.000000	0.000000	0.000000	0.000368	0.000000	0.000000	0.000000
щ	0.000557	0.000003	0.000010	0.000000	0.000003	0.001626	0.000000	0.000000	0.001348	0.000000	0.000000	0.000000	0.000000	0.000058	0.000006	0.000003	0.000000	0.000003	0.000006	0.000179	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000029	0.000000	0.000000	0.000000
ы	0.000093	0.000467	0.001489	0.000227	0.000256	0.001233	0.000125	0.000247	0.000381	0.002199	0.000435	0.002225	0.001726	0.000650	0.000256	0.000618	0.000391	0.001210	0.000912	0.000144	0.000016	0.001127	0.000029	0.000298	0.000595	0.000006	0.000000	0.000000	0.000048	0.000000	0.000061	
ь	0.000189	0.000391	0.001287	0.000291	0.000480	0.001245	0.000035	0.000515	0.000986	0.000000	0.001726	0.000166	0.000695	0.002708	0.000698	0.001197	0.000218	0.001844	0.000816	0.000243	0.000045	0.000074	0.000147	0.000544	0.000423	0.000016	0.000000	0.000000	0.000157	0.000384	0.000573	
э	0.000000	0.000000	0.000003	0.000006	0.000000	0.000000	0.000000	0.000000	0.000006	0.000000	0.000000	0.000042	0.000048	0.000006	0.000026	0.000000	0.000013	0.000003	0.000026	0.002801	0.000003	0.000006	0.000029	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
ю	0.000090	0.000259	0.000336	0.000128</																												

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

$$H_0 = \log_2 32 = 5$$

Без пробіла, H_1 :	$R = 1 - \frac{4,452870917974479}{5} = 0,109425816$
З пробілом, H_1 :	$R = 1 - \frac{4,379945305657554}{5} = 0,124010939$
Перетинаються, з пробілом, H_2 :	$R = 1 - \frac{3,991013097121348}{5} = 0,201797381$
Не перетинаються, з пробілом, H_2 :	$R = 1 - \frac{3,990569680311937}{5} = 0,201886064$
Перетинаються, без пробіла, H_2 :	$R = 1 - \frac{4,148671405718143}{5} = 0,170265719$
Не перетинаються, без пробіла, H_2 :	$R = 1 - \frac{4,147853620317564}{5} = 0,170429276$

Код програми

lab1_symbol_entropy.py:

```
import math
#symbol entropy without space
f = open('text.txt','r',encoding = 'utf-8')
mytext = f.read().replace('-', '').replace('?', '').replace('""', '').replace(' ', '').replace('\n',
 '').lower().replace('!', '').replace('(', '').replace(')', '').replace('.', '').replace(',', '').replace(':', '').replace(';','')
f.close()
length = len(mytext)
frequency_dictionary = { } #dictionary of symbols(=keys) and their counts
for symbol in mytext:
    count = frequency_dictionary.get(symbol,0)
    frequency_dictionary[symbol] = count + 1

frequency_symbols = frequency_dictionary.keys() #symbols(=keys), набор ключей словаря
entropy = 0
for symbols in frequency_symbols:
    print (symbols, frequency_dictionary[symbols]/length)
    entropy -= (frequency_dictionary[symbols]/length)*math.log(frequency_dictionary[symbols]/length,2)
print (entropy)
print('-----')

#symbol entropy with space
f2 = open('text.txt','r',encoding = 'utf-8')
mytext2 = f2.read().replace('-', '').replace('?', '').replace('""', '').replace('\n',
 '').lower().replace('!', '').replace('(', '').replace(')', '').replace('.', '').replace(',', '').replace(':', '').replace(';','')
f2.close()
length2 = len(mytext2)
frequency_dictionary2 = { }
for symbol in mytext2:
    count2 = frequency_dictionary2.get(symbol,0)
    frequency_dictionary2[symbol] = count2 + 1

frequency_symbols2 = frequency_dictionary2.keys()
entropy2 = 0
for symbols in frequency_symbols2:
    print (symbols, frequency_dictionary2[symbols]/length2)
    entropy2 -= (frequency_dictionary2[symbols]/length2)*math.log(frequency_dictionary2[symbols]/
length2,2)
print (entropy2)
```

lab1_bigrams without spaces.py:

```
import re
import math
frequency = {}
f = open('text.txt', 'r',encoding = 'utf-8')
mytext = f.read().replace('?', '').replace(' ', '').replace('""', '').replace('\n',
 '').lower().replace('!', '').replace('(', '').replace(')', '').replace('.', '').replace(',', '').replace(':', '').replace(';','')
f.close()
match_pattern = re.findall('\w\w', mytext)
```

```

for bigram in match_pattern:
    count = frequency.get(bigram,0)
    frequency[bigram] = count + 1

frequency_list = frequency.keys()
length = len(mytext)
entropy = 0
for bigrams in frequency_list:
    print (bigrams, frequency[bigrams]/length)
    entropy -= (frequency[bigrams]/length) * math.log(frequency[bigrams]/length,2)
print (entropy)
print('-----')

```

lab1_bigrams with spaces.py:

```

import re
import math
frequency = {}
f = open('text.txt', 'r', encoding = 'utf-8')
mytext = f.read().replace('?', '').replace('""', '').replace('-', '').replace('\n',
 '').lower().replace('!', '').replace('(', '').replace(')', '').replace('.', '').replace(',', '').replace(':', '').replace('; ', '')
f.close()
match_pattern = re.findall('\w.|\.\w', mytext)

for bigram in match_pattern:
    count = frequency.get(bigram,0)
    frequency[bigram] = count + 1

frequency_list = frequency.keys()
length = len(mytext)
entropy = 0
for bigrams in frequency_list:
    print (bigrams, frequency[bigrams]/length)
    entropy -= (frequency[bigrams]/length) * math.log(frequency[bigrams]/length,2)
print (entropy)
print('-----')

```

lab1_alteration bigrams without spaces.py

```

import re
import math

f = open('text.txt', 'r', encoding = 'utf-8')
mytext = f.read().replace('-', '').replace('?', '').replace(' ', '').replace('""', '').replace('\n',
 '').lower().replace('!', '').replace('(', '').replace(')', '').replace('.', '').replace(',', '').replace(':', '').replace('; ', '')
f.close()

output = open("text_output.txt", "r+")
for i in range(1, len(mytext)):
    output.write(mytext[i-1])
    output.write(mytext[i])

```

```

output.close()

frequency = {}
f1 = open('text_output.txt', 'r')
mytext = f1.read()
f1.close()
match_pattern = re.findall('\w\w', mytext)

for bigram in match_pattern:
    count = frequency.get(bigram,0)
    frequency[bigram] = count + 1

frequency_list = frequency.keys()
lenght = len(mytext)
enthropy = 0
for bigrams in frequency_list:
    print (bigrams, frequency[bigrams]/lenght)
    enthropy -= (frequency[bigrams]/lenght)*math.log(frequency[bigrams]/lenght,2)
print (enthropy)
print('-----')

```

lab1_alteration bigrams with spaces.py

```

import re
import math

f = open('text.txt', 'r', encoding = 'utf-8')
mytext = f.read().replace('-', '').replace('?', '').replace('""', '').replace('\n',
 '').lower().replace('!', '').replace('(', '').replace(')', '').replace('.', '').replace(',', '').replace(':', '').replace('; ', '')
f.close()

output = open("text_output2.txt", "r+")
for i in range(1,len(mytext)):
    output.write(mytext[i-1])
    output.write(mytext[i])
output.close()

frequency = {}
f1 = open('text_output2.txt', 'r')
mytext = f1.read()
f1.close()
match_pattern = re.findall('\w.|\.\w', mytext)

for bigram in match_pattern:
    count = frequency.get(bigram,0)
    frequency[bigram] = count + 1

frequency_list = frequency.keys()
lenght = len(mytext)
enthropy = 0
for bigrams in frequency_list:
    print (bigrams, frequency[bigrams]/lenght)

```

```
    entropy -= (frequency[bigrams]/length) * math.log(frequency[bigrams]/length,2)
print (entropy)
print('-----')
```