

Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Фізико-технічний інститут

Симетрична криптографія

Комп'ютерний практикум №1
Експериментальна оцінка ентропії на символ джерела відкритого тексту

Варіант 9

Виконала:
студентка групи ФІ-93
Ліщинська О.Т.

Перевірив:
Чорний О.М.

Київ – 2022

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

0. Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.
1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку $H(1)$ та $H(2)$ за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення $H(1)$ та $H(2)$ на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення $H(1)$ та $H(2)$ на тому ж тексті, в якому вилучено всі пробіли.
2. За допомогою програми CoolPinkProgram оцінити значення $H(10)$, $H(20)$, $H(30)$.
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

В даному практикумі найбільшою проблемою для мене був вибір способу обрахунку частот біграм з перетином і без і його реалізація, оскільки від вибору варіанту рішення залежала б оптимальність моєї програми.

Я вирішила, що найкращим варіантом буде «пробігати» текст лише один раз і одразу додавати в масиві значення, що відповідають тим парам символів, які мені попадалися. Тобто, проходячи текст, я знаходила пару і одразу додавала +1 до значення, яке знаходилось в матриці на масці, що є перетином відповідних знайдених символів.

Таблиці частот букв і біграм

(для тексту без пробілів)

Частоти літер

А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П	Р
0.0774438	0.0189546	0.0459994	0.0177162	0.0308197	0.0852628	0.0110476	0.0165808	0.0687849	0.0104397	0.0420787	0.0451389	0.0300602	0.0626962	0.113404	0.0279651	0.0414595

С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я	Ё
0.0504701	0.0603336	0.0300939	0.00106436	0.00952686	0.00355971	0.0190799	0.0101311	0.00290501	0.000301163	0.0189471	0.0201761	0.00214929	0.00630946	0.0189639	0.00436967

Частоти біграм без перетинів

	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П	Р
А	0.000179574	0.00153762	0.0054359	0.0011635	0.00255895	0.00184065	0.00190425	0.00470638	0.00124581	0.000931548	0.0102059	0.00866078	0.00474379	0.00576138	0.000804349	0.0027722	0.00310142
Б		0	0	0	0	0.00269363	0	0	0.000662185	0	0.000187058	0.000931548	0	0.000374116	0.00260385	0	0.00178827
В	0.00628888	0.000276493	0.000443328	0.000405916	0.000858596	0.00639738	0	0.000665927	0.0042219	0	0.000997019	0.000815573	0.000381598	0.00198281	0.00786766	0.00106997	0.0010681
Г	0.000980183	0	0.000153387	0	0.00120839	0.00042649	0	0	0.000774419	0	0.000115976	0.00144409	0	0.000396563	0.00991032	0	0.00081183
Д	0.00595592	0	0.00119717	0	0	0.00557806	0	0	0.00261133	0	0.000407786	0.000695855	0.000149646	0.00226714	0.004018	0.00022821	0.00172841
Е	0.000280587	0.00274975	0.00466148	0.00411527	0.00456047	0.0017546	0.00119717	0.00211001	0.00144409	0.00224844	0.00272356	0.00742246	0.00568282	0.00973449	0.00142912	0.00384217	0.00692114
Ж	0.00110364	0.000104752	0	0	0.000886654	0.00489717	0	0	0.0016087	0	0.000254399	0	0	0.00121213	0	0	0
З	0.00577261	0.000273104	0.00125703	0.000362892	0.00119343	0.00052376	0.00019454	0	0.000400304	0	0.000280587	0.000243175	0.000198281	0.00218109	0.000793125	0.00017583	0.00019454
И	0.000216987	0.00191921	0.00509546	0.00115976	0.00254025	0.00306775	0.00052002	0.00307897	0.00173216	0.0014815	0.00485602	0.00540971	0.00384965	0.0060906	0.00142164	0.00306401	0.00115602
Й	0.000288065	0.000370375	0.000523762	0.000276846	0.000669667	0.00011223	0.00011971	0.000198281	0.000505056	0	0.000935289	0.000142164	0.000482609	0.0010999	0.000366633	0.00077816	0.00029555
К	0.0108157	0.000437715	0.000875431	0.000164611	0.000288069	0.00092406	0.00017957	0.000134682	0.00497574	0	0.000415268	0.000722043	0.000179576	0.00143286	0.0127573	0.00054995	0.00190799
Л	0.0052264	0.00041901	0.000766937	0.000366633	0.000325481	0.00496452	0.00038159	0.000198281	0.00814824	0	0.000823055	0.000198281	0.000149646	0.00104752	0.00772549	0.00075571	0.00010475
М	0.00303408	0.000374116	0.000852984	0.000306775	0.000703337	0.00444075	0.00013842	0.00025814	0.00435471	0	0.000587362	0.000299293	0.000280587	0.00252154	0.00392447	0.00108868	0.00022072
Н	0.0109578	0.000310516	0.000329222	0.000119717	0.000385339	0.0112609	0	0	0.00915461	0	0.000692114	0	0	0.00377483	0.0119081	0.00055743	0.00020202
О	0.00025814	0.00588858	0.0124543	0.00595592	0.0070446	0.0027909	0.0027909	0.00250658	0.00226714	0.00356906	0.00361396	0.00692114	0.00695481	0.00857847	0.00183691	0.00400304	0.00802478
П	0.0014366	0	0	0	0	0.00240556	0	0	0.0012271	0	0.000220728	0.000755714	0	0.00013094	0.0114442	0	0.00858221
Р	0.00738878	0.00013094	0.000344186	0.00019454	0.000366633	0.00674905	0.00029929	0	0.0056903	0	0.00051628	0.000101011	0.000202022	0.000976442	0.00862337	0.00019828	0
С	0.00213246	0.000209505	0.00180324	0	0.000523762	0.00472508	0	0.000115976	0.00158625	0	0.00472882	0.00329596	0.00109242	0.00132811	0.00339697	0.00189303	0.00023195
Т	0.00696603	0.000385339	0.0030266	0.000179576	0.000392821	0.00632256	0.00025814	0.000190799	0.00410405	0	0.00110738	0.000325481	0.000231952	0.000232326	0.0164798	0.00070333	0.00275349
У	0.00016461	0.00127199	0.00150769	0.00172841	0.00290314	0.00037785	0.00240556	0.000624773	0.000695855	0.000127199	0.00144783	0.00151517	0.00157129	0.00130566	0.000336704	0.0018182	0.00083801
Ф	0.00028432	0	0	0	0	0.00018331	0	0	0.000213246	0	0	0	0	0	0	0	0.00013468
Ч	0.00096896	0.000190799	0.000591103	0.00016087	0.000284328	0.00019454	0	0.000115976	0.000677149	0	0.000314257	0.000321739	0.000213246	0.000609809	0.00284328	0.00056491	0.00027310
Ц	0.000852984	0	0.000145905	0	0	0.00087543	0	0	0.000213246	0	0	0	0	0	0.000284328	0	0
Ш	0.0024093	0	0	0	0	0.0042537	0	0	0.00410779	0	0.00058362	0	0	0.00105501	0.00013094	0.00010101	0
Щ	0.00116724	0	0	0	0	0.00257017	0	0	0.00306027	0	0.000998889	0.000520021	0	0.000269363	0.000362892	0	0
Ъ	0.00039282	0	0	0	0	0.00153013	0	0	0.000882913	0	0	0	0	0	0	0	0
Ы	0	0	0	0	0	0.00014216	0	0	0	0	0	0	0	0	0	0	0
Ы	0	0.000404045	0.00148898	0.000243175	0.000377857	0.00172841	0	0.000205764	0.000441457	0.00191173	0.000703337	0.00281709	0.00159747	0.000879172	0.000288069	0.00075197	0.00043771
Ь	0.000190799	0.000591103	0.00131315	0.000299293	0.000718302	0.00099140	0	0.000557432	0.00101011	0	0.0024542	0.000187058	0.000688373	0.00266744	0.000699596	0.00121962	0.00023195
Э	0	0	0	0	0	0	0	0	0	0	0.000145905	0	0	0	0	0	0
Ю	0	0.000456421	0.00025814	0.000115976	0.000568656	0	0	0.000119717	0.000299293	0	0.000430233	0	0.000157129	0.000430233	0.000202022	0.00043397	0.00019079
Я	0.000190799	0.000520021	0.00151891	0.000299293	0.00112983	0.00031051	0.00021324	0.000594844	0.000864207	0.000202022	0.00128696	0.000781902	0.000860466	0.00235319	0.000617291	0.00114854	0.00043771
Ё	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я	Ё
А	0.00533489	0.00621032	0.000512538	0.000112235	0.00129444	0.000119717	0.00172093	0.0010363	0.000254399	0	0	0	0.000273104	0.000834278	0.00216613	0
Б	0.00018331	0	0.00177331	0	0	0	0	0	0.000123458	0	0.0049907	0	0	0	0.000602326	0
В	0.00451184	0.000759456	0.000858596	0	0.000112235	0	0.00029181	0.00179389	0	0	0.00308459	0.000220728	0.000188929	0	0.000364763	0
Г	0.00012719	0	0.000931548	0	0	0	0	0	0	0	0	0	0	0	0	0
Д	0.00070333	0.000314257	0.0021362	0	0	0.000273104	0.00011223	0	0	0	0.000531244	0.00106997	0	0	0.000370375	0
Е	0.00747109	0.00707079	0.000707079	0	0.00115976	0.000755714	0.00233448	0.0010999	0.000987665	0	0	0	0.000243175	0.000265622	0.000314257	0
Ж	0.00010101	0	0.000202022	0	0	0	0	0	0	0	0	0	0	0	0	0
З	0.00017957	0.000119717	0.000325481	0	0	0	0.00013094	0	0	0	0.000467645	0.000153387	0	0	0.000688373	0
И	0.00474753	0.00543964	0.000546209	0.000254399	0.00227836	0.00132437	0.00392447	0.000755714	0.000250657	0	0	0	0.000157129	0.000374116	0.00164611	0
Й	0.0014179	0.000617291	0.000183317	0	0	0	0.00041526	0.000224465	0	0	0	0	0	0	0	0
К	0.00062851	0.000953995	0.00307523	0	0	0	0.00040778	0	0	0	0	0	0	0	0	0
Л	0.00245794	0.00038908	0.00184065	0	0	0	0.00075571	0	0	0	0.000703337	0.00489717	0	0.000998889	0.00170971	0
М	0.00106997	0.000433974	0.00274975	0	0.000112235	0	0.00047886	0	0	0	0.00100637	0	0	0	0.00058362	0
Н	0.00083053	0.000490092	0.00317624	0	0	0.000284328	0.00035166	0	0	0	0.00411901	0.00177705	0	0.000209505	0.0012271	0
О	0.00933044	0.00925188	0.00084176	0.000153387	0.000909101	0.000119717	0.00355036	0.00138049	0.000235693	0	0	0	0.000265622	0.00114479	0.000845501	0
П	0	0	0.000733267	0	0	0	0	0	0	0	0.000205764	0	0	0	0.000303034	0
Р	0.00046390	0.00105127	0.00324732	0	0.000179576	0	0	0.000609805	0	0	0.00152639	0.000617291	0	0.000183317	0.00122336	0
С	0.00090536	0.0119118	0.00114854	0	0.000224469	0	0.00034044	0.000142164	0	0	0.000325481	0.00327351	0	0.000216987	0.0047999	0.000127199
Т	0.00184439	0.000456421	0.00192295	0	0	0	0.00047512	0	0	0	0.00174338	0.00716806	0.000101011	0.000101011	0.000673408	0
У	0.00206512	0.00261133	0.000119717	0	0.000725784	0	0.00156754	0.000987665	0.000183317	0	0	0	0	0.00140667	0.000127199	0
Ф	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ч	0.00048260	0.000224469	0.000306775	0	0	0	0.00016087	0	0	0	0	0	0	0	0	0
Ц	0	0	0.000276846	0	0	0	0	0	0	0	0.000299293	0	0	0	0	0
Ш	0	0.00457543	0.000830537	0	0	0	0	0.000213246	0	0	0	0.000202022	0	0	0	0
Щ	0	0	0.000317998	0	0	0	0	0	0	0	0	0.000695855	0	0	0	0
Ъ	0	0	0.000104752	0	0	0	0	0	0	0	0	0	0	0	0	0
Ы	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000145905	0
Ы	0.00123458	0.00104004	0.000153387	0	0.00119343	0	0.00031799	0.000534985	0	0	0	0	0	0	0	0
Ь	0.00215865	0.000890395	0.000269363	0	0.000119717	0.000202022	0.00072952	0.000542468	0	0	0	0	0.000138423	0.000303034	0.000624773	0
Э	0	0.00184813	0	0												

(для тексту з пробілами)

Частоти літер

A	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П	Р
0.0647795	0.015855	0.0384772	0.0148191	0.0257798	0.0713199	0.00924103	0.0138694	0.0575366	0.00873251	0.0351976	0.0377574	0.0251445	0.0524436	0.0948591	0.0233921	0.0346797
С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я	Ё	space
0.0422168	0.0504674	0.0251727	0.000890306	0.00796894	0.0029776	0.0159598	0.00847434	0.00242996	0.000251914	0.0158487	0.0168767	0.00179782	0.00527768	0.0158628	0.00365511	0.163528

Частоти біграм з перетинами

	A	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П	Р
A	0	0.000696286	0.00295569	0.000614922	0.00147081	0.001161	0.00144264	0.00356279	0	0.000740097	0.00759812	0.00693469	0.00350333	0.00341728	0	0.00084023	0.00219682
Б	0.00122515	0	0	0	0	0.002189	0	0	0.000531994	0	0.000129869	0.00079799	0	0.000280079	0.00208416	0	0.00151149
В	0.00512278	0	0	0	0.000308243	0.00526674	0	0.000425595	0.00324985	0	0.000237832	0.000578934	0.000161163	0.00114692	0.00621337	0.00023626	0.000646214
Г	0.000779214	0	0	0	0.00101705	0.00034892	0	0	0.000639957	0	0	0.00111406	0	0.000264432	0.00831162	0	0.000715062
Д	0.00501326	0	0.000857448	0	0	0.00465181	0	0	0.0019496	0	0.00027382	0.000586758	0	0.00175245	0.00325455	0.00011735	0.00145203
Е	0	0.00142543	0.00207634	0.00308713	0.00282739	0.0011563	0.00081363	0.00111406	0.000139257	0.0018964	0.00131277	0.00598649	0.00391172	0.00627126	0.000212797	0.00100766	0.00538252
Ж	0.000932553	0	0	0	0.000677509	0.00413234	0	0	0.00131434	0	0.000184633	0	0	0.000888742	0	0	0
З	0.00499135	0.000179939	0.000987317	0.000256609	0.000956024	0.00042716	0.00014551	0	0.000316067	0	0	0.000175245	0.000154904	0.00169769	0.000560158	0	0.000109524
И	0	0.000802684	0.00220934	0.000521041	0.00131434	0.00212015	0.00026286	0.00204661	0.000361443	0.00119386	0.00298386	0.00446249	0.00265684	0.00332027	0.000125175	0.00023626	0.000627435
Й	0	0	0	0	0.000129869	0	0	0	0	0	0.000256609	0	0	0.00030042	0	0	0
К	0.00905954	0	0.000195586	0	0	0.00073853	0	0	0.00380375	0	0	0.000557029	0	0.000542946	0.0102988	0	0.00147394
Л	0.0043201	0	0	0.000156469	0	0.0036864	0.00028946	0	0.00631664	0	0.000345796	0.000106399	0	0.000187762	0.00586288	0	0
М	0.00243309	0	0	0	0	0.00367388	0	0	0.00304488	0	0	0.000159598	0	0.0012361	0.00277106	0	0
Н	0.00905484	0	0	0	0.00017368	0.00952112	0	0	0.00742287	0	0.000391172	0	0	0.00278045	0.00981998	0	0
О	0	0.0037177	0.00803154	0.00442024	0.00497257	0.00197776	0.00186198	0.00140978	0.000802684	0.00302298	0.002056	0.00545763	0.0051932	0.0047457	0.000183068	0.00124549	0.00590669
П	0.00126896	0	0	0	0	0.00211546	0	0	0.00102643	0	0.000186198	0.000671251	0	0	0.00954146	0	0.00704266
Р	0.0061367	0	0.000220621	0.000131434	0.000262867	0.00564696	0.00022218	0	0.00478951	0	0.000312937	0.000101705	0.00017368	0.000772955	0.00722885	0	0
С	0.00174463	0	0.00138318	0	0.000308243	0.00378028	0	0	0.00121733	0	0.00369892	0.00275072	0.000719756	0.000765132	0.00266623	0.0012361	0.000153335
Т	0.00558437	0	0.00202783	0	0.000128304	0.0050336	0	0	0.00294318	0	0.000561723	0.000148645	0	0.00136284	0.0135095	0.00011578	0.00232043
У	0	0.000796426	0.00070098	0.00127209	0.00204505	0.00021905	0.00188858	0.000364572	0	0.000106399	0.000815202	0.00112345	0.00112814	0.000195586	0	0.00086996	0.000586758
Ф	0.00022688	0	0	0	0	0.00014708	0	0	0.00017681	0	0	0	0	0	0	0	0.00012048
Ч	0.000660298	0	0.000206539	0	0	0	0	0	0.000190892	0	0	0.000203409	0	0.000167422	0.00216866	0	0.000109524
Ц	0.000708803	0	0	0	0	0.00077139	0	0	0.00015021	0	0	0	0	0	0.000206539	0	0
Ч	0.00207634	0	0	0	0	0.00349082	0	0	0.00339537	0	0.000505394	0	0	0.000827719	0	0	0
Ш	0.000941942	0	0	0	0	0.00214988	0	0	0.0023971	0	0.000768261	0.00042716	0	0.00025035	0.000341102	0	0
Щ	0.000297291	0	0	0	0	0.00128617	0	0	0.000696286	0	0	0	0	0	0	0	0
Ъ	0	0	0	0	0	0.00012361	0	0	0	0	0	0	0	0	0	0	0
Ы	0	0.000143951	0.000882483	0.000145516	0	0.00133624	0	0	0	0.00157407	0.000283208	0.00233764	0.00119699	0.000175245	0	0.00018306	0.000305114
Ь	0	0	0	0	0	0.00047566	0	0.000214362	0	0	0.00146142	0	0.00032389	0.00118447	0	0	0
Э	0	0	0	0	0	0	0	0	0	0	0.000109528	0	0	0	0	0	0
Ю	0	0.000233138	0	0	0.000272256	0	0	0	0	0	0	0	0	0	0	0	0
Я	0	0	0.000281644	0	0.000442806	0	0.00012674	0.000206539	0	0.000172116	0.000452195	0.000566417	0.000463147	0.000711933	0	0	0.00013925
Ё	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
space	0.00212797	0.00736185	0.0152048	0.00399621	0.00818801	0.00352524	0.00203096	0.0043467	0.0104318	0	0.00995454	0.00240962	0.00513374	0.0170488	0.00920349	0.0170207	0.00362694

	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я	Ё	space
А	0.002951	0.00447344	0	0	0.00101235	0	0.00087622	0.00077295	0.000248785	0	0	0	0	0.000718191	0.00171333	0	0.0146283
Б	0.00014551	0	0.00154748	0	0	0	0	0	0.000117352	0	0.00416989	0	0	0	0.000485053	0	0.000295726
В	0.00305114	0.000142387	0.000586758	0	0	0	0	0.00144577	0	0	0.00258017	0.000184633	0	0	0.00027382	0	0.00640896
Г	0	0	0.000776085	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000691592
Д	0.00046001	0.000236268	0.00176497	0	0	0.000211233	0	0	0	0	0.0004475	0.000895001	0	0	0.000311373	0	0.0011078
Е	0.00464712	0.00498196	0.000109528	0	0.000733838	0.000586758	0.00113127	0.00084806	0.00080112	0	0	0	0	0.000228444	0.000115787	0	0.018479
Ж	0	0	0.000175245	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000643086
З	0	0	0.000236268	0	0	0	0	0	0	0	0.000392736	0.000136128	0	0	0.00052417	0	0.00132529
И	0.00230322	0.00370361	0	0.000156469	0.00172585	0.00108746	0.00248629	0.00051321	0.000172116	0	0	0	0	0.000294161	0.00129869	0	0.0185259
Й	0.00042559	0.000215927	0	0	0	0	0	0.000117352	0	0	0	0	0	0	0	0	0.00704109
К	0	0.000499135	0.00241275	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00548266
Л	0.0014223	0	0.00133155	0	0	0	0	0	0	0	0.000585193	0.00401968	0	0.000843366	0.00142074	0	0.00652944
М	0	0	0.00217961	0	0	0	0	0	0	0	0.000816767	0	0	0.000449065	0.000795956	0	0.00795956
Н	0.00033640	0.000286338	0.00262085	0	0	0.000239397	0.00016116	0	0	0	0.0033453	0.00138944	0	0.000175245	0.000960718	0	0.00357218
О	0.00567668	0.00628378	0	0	0.000502264	0	0.00204661	0.00111875	0.000153339	0	0	0	0	0.000999835	0.000528864	0	0.0222874
П	0	0	0.000632134	0	0	0	0	0	0	0	0.000179939	0	0	0	0.000283208	0	0
Р	0.00028164	0.000843366	0.00273977	0	0.000159598	0	0	0.00047253	0	0	0.00135971	0.000542946	0	0.000156469	0.000987317	0	0.000791732
С	0.00065091	0.00983249	0.000782343	0	0.000165857	0	0.00023783	0	0	0	0.00030042	0.00266466	0	0.000172116	0.00402281	0	0.0026099
Т	0.00117508	0	0.00153809	0	0	0.00015333	0	0	0	0	0.00142387	0.00602404	0	0	0.000467841	0	0.00559219
У	0.00113283	0.00163197	0	0	0.000539817	0	0.00089500	0.00083554	0.000168986	0	0	0	0	0.00112188	0	0	0.00664679
Ф	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ч	0	0	0.000172116	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00389451
Ц	0	0	0.000215927	0	0	0	0	0	0	0	0.000239397	0	0	0	0	0	0.000564852
Ш	0	0.00388042	0.000718191	0	0	0	0	0.000154904	0	0	0	0.000187762	0	0	0	0	0.000527299
Щ	0	0	0.000289467	0	0	0	0	0	0	0	0	0.000569546	0	0	0	0	0.000165857
Ъ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ы	0.00057111	0.000608663	0	0	0.000995141	0	0.00012986	0.00044280	0	0	0	0	0	0	0	0	0.00429507
Ь	0.00097010	0.000131434	0	0	0	0.000131434	0	0.00040994	0	0	0	0	0	0.000264432	0.000452195	0	0.0105163
Э	0	0.00149897	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ю	0.00012674	0.000361443	0	0	0	0	0.00011422	0.00015333	0.000303549	0	0	0	0	0	0	0	0.00340163
Я	0.00037552	0.00153496	0	0	0.000209668	0	0.00014708	0	0.000134563	0	0	0	0	0	0	0	0.00937717
Ё	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
space	0.0151618	0.00896566	0.00407288	0.000449065	0.00169299	0.00032389	0.00705048	0.00088717	0.000181504	0	0	0	0.00175871	0	0.00132842	0	0

Частоти біграм без перетинів

	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П	Р
А	0	0.000696286	0.00295569	0.000614922	0.00147081	0.001161	0.00144264	0.00356279	6.72815e-005	0.000740097	0.00759812	0.00693469	0.00350333	0.00341728	0	0.00084023	0.00219682
Б	0.00122515	0	0	0	0	0.002189	0	0	0.000531994	0	0.000129869	0.00079799	0	0.000280079	0.00208416	0	0.00151149
В	0.00512278	0	0	0	0.000308243	0.00526674	0	0.000425595	0.00324985	0	0.000237832	0.000578934	0.000161163	0.00114692	0.00621337	0.00023626	0.00064621
Г	0.000779214	0	0	0	0.00101705	0.00034892	0	0	0.000639957	0	0	0.00111406	0	0.000264432	0.00831162	0	0.00071506
Д	0.00501326	0	0.000857448	0	0	0.00465181	0	0	0.0019496	0	0.00027382	0.000586758	0	0.00175245	0.00325455	0.00011735	0.00145203
Е	0	0.00142543	0.00207634	0.00308713	0.00282739	0.0011563	0.00081363	0.00111406	0.000139257	0.0018964	0.00131277	0.00598649	0.00391172	0.00627126	0.000212797	0.00100766	0.00538252
Ж	0.000932553	0	0	0	0.000677509	0.00413234	0	0	0.00131434	0	0.000184633	0	0	0.000888742	0	0	0
З	0.00499135	0.000179939	0.000987317	0.000256609	0.000956024	0.00042716	0.00014551	0	0.000316067	0	0	0.000175245	0.000154904	0.00169769	0.000560158	0	0.00010952
И	0	0.000802684	0.00220934	0.000521041	0.00131434	0.00212015	0.00026286	0.00204661	0.000361443	0.00119386	0.00298386	0.00446249	0.00265684	0.00332027	0.000125175	0.00023626	0.00062743
Й	0	0	0	0	0.000129869	0	0	0	0	0	0.000256609	0	0	0.00030042	0	0	0
К	0.00905954	0	0.000195586	0	0	0.00073853	0	0	0.00380375	0	0	0.000557029	0	0.000542946	0.0102988	0	0.00147394
Л	0.0043201	0	0	0.000156469	0	0.0036864	0.00028946	0	0.00631664	0	0.000345796	0.000106399	0	0.000187762	0.00586288	0	0
М	0.00243309	0	0	0	0	0.00367388	0	0	0.00304488	0	0	0.000159598	0	0.0012361	0.00277106	0	0
Н	0.00905484	0	0	0	0.00017368	0.00952112	0	0	0.00742287	0	0.000391172	0	0	0.00278045	0.00981998	0	0
О	0	0.0037177	0.00803154	0.00442024	0.00497257	0.00197776	0.00186198	0.00140978	0.000802684	0.00302298	0.002056	0.00545763	0.0051932	0.0047457	0.000183068	0.00124549	0.00590669
П	0.00126896	0	0	0	0	0.00211546	0	0	0.00102643	0	0.000186198	0.000671251	0	0	0.00954146	0	0.00704266
Р	0.0061367	0	0.000220621	0.000131434	0.000262867	0.00564696	0.00022218	0	0.00478951	0	0.000312937	0.000101705	0.00017368	0.000772955	0.00722885	0	0
С	0.00174463	0	0.00138318	0	0.000308243	0.00378028	0	0	0.00121733	0	0.00369892	0.00275072	0.000719756	0.000765132	0.00266623	0.0012361	0.00015333
Т	0.00558437	0	0.00202783	0	0.000128304	0.0050336	0	0	0.00294318	0	0.000561723	0.000148645	0	0.00136284	0.0135095	0.00011578	0.00232043
У	0	0.000796426	0.00070098	0.00127209	0.00204505	0.00021905	0.00188858	0.000364572	0	0.000106399	0.000815202	0.00112345	0.00112814	0.000195586	0	0.00086996	0.00058675
Ф	0.00022688	0	0	0	0	0.00014708	0	0	0.00017681	0	0	0	0	0	0	0	0.00012048
Ч	0.000660298	0	0.000206539	0	0	0	0	0	0.000190892	0	0	0.000203409	0	0.000167422	0.00216866	0	0.00010952
Ц	0.000708803	0	0	0	0	0.00077139	0	0	0.00015021	0	0	0	0	0	0.000206539	0	0
Ш	0.00207634	0	0	0	0	0.00349082	0	0	0.00339537	0	0.000505394	0	0	0.000827719	0	0	0
Щ	0.000941944	0	0	0	0	0.00214988	0	0	0.0023971	0	0.000768261	0.00042716	0	0.00025035	0.000341102	0	0
Ъ	0.00029729	0	0	0	0	0.00128617	0	0	0.000696286	0	0	0	0	0	0	0	0
Ы	0	0	0	0	0	0.00012361	0	0	0	0	0	0	0	0	0	0	0
Ь	0	0.000143951	0.000882483	0.000145516	0	0.00133624	0	0	0	0.00157407	0.000283208	0.00233764	0.00119699	0.000175245	0	0.00018306	0.00030511
Э	0	0	0	0	0	0.00047566	0	0.000214362	0	0	0	0.00146142	0	0.00032389	0.00118447	0	0
Ю	0	0	0	0	0	0	0	0	0	0	0.000109528	0	0	0	0	0	0
Я	0	0	0.000281644	0	0.000442806	0	0.00012674	0.000206539	0	0.000172116	0.000452195	0.000566417	0.000463147	0.000711933	0	0	0.00013925
Ё	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
space	0.00212797	0.00736185	0.0152048	0.00399621	0.00818801	0.00352524	0.00203096	0.0043467	0.0104318	0	0.00995454	0.00240962	0.00513374	0.0170488	0.00920349	0.0170207	0.00362694

	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я	Ё	space
А	0.002951	0.00447344	0	0	0.00101235	0	0.00087622	0.00077295	0.000248785	0	0	0	0	0.000718191	0.00171333	0	0.0146283
Б	0.00014551	0	0.00154748	0	0	0	0	0	0.000117352	0	0.00416989	0	0	0	0.000485053	0	0.000295726
В	0.00305114	0.000142387	0.000586758	0	0	0	0	0.00144577	0	0	0.00258017	0.000184633	0	0	0.00027382	0	0.00640896
Г	0	0	0.000776085	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000691592
Д	0.00046001	0.000236268	0.00176497	0	0	0.000211233	0	0	0	0	0.0004475	0.000895001	0	0	0.000311373	0	0.0011078
Е	0.00464712	0.00498196	0.000109528	0	0.000733838	0.000586758	0.00113127	0.00084806	0.00080112	0	0	0	0	0.000228444	0.000115787	0	0.018479
Ж	0	0	0.000175245	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000643086
З	0	0	0.000236268	0	0	0	0	0	0	0	0.000392736	0.000136128	0	0	0.00052417	0	0.00132529
И	0.00230322	0.00370361	0	0.000156469	0.00172585	0.00108746	0.00248629	0.00051321	0.000172116	0	0	0	0	0.000294161	0.00129869	0	0.0185259
Й	0.00042559	0.000215927	0	0	0	0	0	0.000117352	0	0	0	0	0	0	0	0	0.00704109
К	0	0.000499135	0.00241275	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00548266
Л	0.0014223	0	0.00133155	0	0	0	0	0	0	0	0.000585193	0.00401968	0	0.000843366	0.00142074	0	0.00652944
М	0	0	0.00217961	0	0	0	0	0	0	0	0.000816767	0	0	0	0.000449065	0	0.00795956
Н	0.00033640	0.000286338	0.00262085	0	0	0.000239397	0.00016116	0	0	0	0.0033453	0.00138944	0	0.000175245	0.000960718	0	0.00357218
О	0.00567668	0.00628378	0	0	0.000502264	0	0.00204661	0.00111875	0.000153339	0	0	0	0	0.00099835	0.000528864	0	0.0222874
П	0	0	0.000632134	0	0	0	0	0	0	0	0.000179939	0	0	0	0.000283208	0	0
Р	0.00028164	0.000843366	0.00273977	0	0.000159598	0	0	0.000472539	0	0	0.00135971	0.000542946	0	0.000156469	0.000987317	0	0.000791732
С	0.00065091	0.00983249	0.000782343	0	0.000165857	0	0.00023783	0	0	0	0.00030042	0.00266466	0	0.000172116	0.00402281	0	0.0026099
Т	0.00117508	0	0.00153809	0	0	0	0.00015333	0	0	0	0.00142387	0.00602404	0	0	0.000467841	0	0.00559219
У	0.00113283	0.00163197	0	0	0.000539817	0	0.00089500	0.00083554	0.000168986	0	0	0	0	0.00112188	0	0	0.00664679
Ф	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ч	0	0	0.000172116	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00389451
Ц	0	0	0.000215927	0	0	0	0	0	0	0	0.000239397	0	0	0	0	0	0.000564852
Ш	0	0.00388042	0.000718191	0	0	0	0	0.000154904	0	0	0	0.000187762	0	0	0	0	0.000527299
Щ	0	0	0.000289467	0	0	0	0	0	0	0	0	0.000569546	0	0	0	0	0.000165857
Ъ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ы	0.00057111	0.000608663	0	0	0.000995141	0	0.00012986	0.000442806	0	0	0	0	0	0	0	0	0.00429507
Ь	0.00097010	0.000131434	0	0	0	0.000131434	0	0.000409948	0	0	0	0	0	0.000264432	0.000452195	0	0.0105163
Э	0	0.00149897	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ю	0.00012674	0.000361443	0	0	0	0	0.00011422	0.00015333	0.000303549	0	0	0	0	0	0	0	0.00340163
Я	0.00037552	0.00153496	0	0	0.000209668	0	0.00014708	0	0.000134563	0	0	0	0	0	0	0	0.00937717
Ё	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
space	0.0151618	0.00896566	0.00407288	0.000449065	0.00169299	0.00032389	0.00705048	0.00088717	0.000181504	0	0	0	0.00175871	0	0.00132842	0	0

Результати роботи програми

Тут наведені значення $H(1)$ та $H(2)$ для набору окремих символів (літер) та біграм

```

C:\Qt_New\Tools\QtCreator\bin\qtcreator_process_stub.exe
entr_letters_without_space: 4.50229
entr_crossed_bigrams_without_space: 4.1343
entr_n_crossed_bigrams_without_space: 4.13228
entr_letters_with_space: 4.40963
entr_crossed_bigrams_with_space: 3.95866
entr_n_crossed_bigrams_with_space: 3.95894

```

The screenshot shows a software interface for a laboratory experiment on entropy. The interface is divided into several sections:

- Top Section:** Contains a text input field with the text "существовало_определенного_соглашения_по_поводу_правил_игры_в_футбол_этот_з", a field for "Использованные буквы:" (empty), and a list of "Порядок n-граммы:" (5, 10, 15, 20, 25, 30, 35, 40, 45, 50) with "10" selected.
- Input Section:** Includes a "Введенный символ:" field with "a", a "Символ по счету:" field with "1", and a "Номер эксперимента:" field with "50". Below these is a "Поле ввода символов:" field with "a" and two buttons: "Продолжить" and "Другой".
- Entropy Calculation Section:** Displays the inequality "1,7723407081821 < H < 2,37225447733486" under the heading "Неравенство для энтропии:".
- Probability Table Section:** Titled "Двоичная таблица угаданных символов:", it shows a list of 32 probabilities, all set to 0,04.
- Status Bar:** At the bottom, it says "Строка состояния: Вы угадали. Для продолжения опыта нажмите 'Продолжить', или 'Другой' для выбора другого порядка".

Произвольная часть текста:	
<u>д_собой_твердого_основания_потому_что_в_разные_века_различные_цивилизации_n</u>	
Использованные буквы:	
т, п, й, ц, у, к, е,	

Порядок n-граммы:	Введенный символ:	Неравенство для энтропии:
5 символов	<u>н</u>	<u>$1,77061613239524 < H < 2,52462408236634$</u>
10 символов	Символ по счету:	Двоичная таблица угаданных символов:
15 символов	<u>8</u>	
<u>20 символов</u>	Номер эксперимента:	<p>The screenshot shows a binary table titled "Двоичная таблица угаданных символов:". It contains two columns of binary strings representing probabilities for different symbols. The first column has 10 rows, and the second column has 10 rows. The symbols being guessed are listed at the bottom: т, п, й, ц, у, к, е.</p>
25 символов	Номер эксперимента:	
30 символов	Номер эксперимента:	
35 символов	Номер эксперимента:	
40 символов	Поле ввода символов:	
45 символов	<u>н</u>	
50 символов	Продолжить	Другой

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

Произвольная часть текста:

обы_просить_освободить_место_того_кто_занял_его_первым_или_что_ему_дали_кус

Использованные буквы:

д, м, й, ц, у, к, е, н, г, ш, щ, э, х, ф, ы, в, а, п, р, о, л, ж, з, я, ч, с, и,

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:

т

Символ по счету:

28

Номер эксперимента:

50

Неравенство для энтропии:

1,58911308419035 < H < 2,30752902292088

Двоичная таблица угаданных символов:

00000001000000000000000000000000

10000000000000000000000000000000

10000000000000000000000000000000

10000000000000000000000000000000

10000000000000000000000000000000

10000000000000000000000000000000

Поле ввода символов:

т

Продолжить

Другой

Вероятности:

q[1] = 0,6

q[2] = 0,1

q[3] = 0,04

q[4] = 0

q[5] = 0,02

q[6] = 0,04

q[7] = 0,02

q[8] = 0,02

q[9] = 0

q[10] = 0

q[11] = 0,02

q[12] = 0,02

q[13] = 0

q[14] = 0

q[15] = 0,04

q[16] = 0,02

q[17] = 0

q[18] = 0

q[19] = 0

q[20] = 0

q[21] = 0,02

q[22] = 0

q[23] = 0

q[24] = 0

q[25] = 0

q[26] = 0

q[27] = 0

q[28] = 0,04

q[29] = 0

q[30] = 0

q[31] = 0

q[32] = 0

Строка состояния:

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

Розрахунок надлишковості

$$R = 1 - \frac{\text{ентропія для літер тексту без пробілів}}{5} = 1 - \frac{4.50229}{5,0444} = 0,107$$

$$R = 1 - \frac{\text{ентропія для біграм з перетином тексту без пробілів}}{5} = 1 - \frac{4.1343}{5,0444} = 0,18$$

$$R = 1 - \frac{\text{ентропія для біграм без перетинів тексту без пробілів}}{5} = 1 - \frac{4.13228}{5,0444} = 0,18$$

$$R = 1 - \frac{\text{ентропія для літер тексту з пробілами}}{5} = 1 - \frac{4.40963}{5,08746} = 0,13$$

$$R = 1 - \frac{\text{ентропія для біграм з перетином тексту з пробілами}}{5} = 1 - \frac{3,95866}{5,08746} = 0,22$$

$$R = 1 - \frac{\text{ентропія для біграм без перетину тексту з пробілами}}{5} = 1 - \frac{3,95894}{5,08746} = 0,22$$

Висновок:

В ході даного лабораторного практикуму я дослідила ентропію тексту використовуючи алфавіт російської мови враховуючи пробіл та не враховуючи його. На прикладах я побачила, що пробіли не впливають значним чином на значення ентропії як для одиночних символів так і для біграм.

Крім того, я обрахувала надлишковість російської мови і отримала середнє значення 0,2 враховуючи ентропію для біграм і 0,1 враховуючи ентропію для одиночних символів.

