

КРИПТОГРАФІЯ

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

ФБ-23 Невмержицька Дар'я

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.

Хід роботи:

Текст завантажую з <https://www.rulit.me/books/arhivarius-get-386934.html>. Фільтрую його у функції `analyze_text(text)`. Частота рахується шляхом ділення кількості елемента/елементів у тексті на загальну кількість зустрічей цих символів.

Таблиці отриманих частот:

Частота букв у тексті:

Символ	Кількість	Частота
о	5688	0.10978
е	4459	0.08606
а	4204	0.08114
и	3460	0.06678
н	3457	0.06672
т	3139	0.06058
с	2813	0.05429
в	2427	0.04684
р	2387	0.04607
л	2080	0.04014
к	1791	0.03457
м	1731	0.03341
д	1652	0.03188
у	1534	0.02961
п	1505	0.02905
я	1151	0.02221
ы	924	0.01783
ь	898	0.01733
г	895	0.01727
э	888	0.01714
б	812	0.01567
ч	750	0.01447
й	644	0.01243
ж	558	0.01077
ю	443	0.00855
ш	423	0.00816
х	407	0.00786
ц	227	0.00438
щ	199	0.00384
э	134	0.00259
ф	82	0.00158
ё	34	0.00066
ъ	18	0.00035

Частота біграм в тексті з перетинами:

Символ	Кількість	Частота
то	755	0.01457
на	653	0.01260
но	637	0.01229
ст	584	0.01127
по	572	0.01104
ен	549	0.01060
ко	528	0.01019
ос	525	0.01013
ни	515	0.00994
ов	511	0.00986
ра	493	0.00951
ро	488	0.00942
не	483	0.00932
ет	456	0.00880
го	455	0.00878
пр	448	0.00865
ом	442	0.00853
ер	430	0.00830
ва	427	0.00824
от	427	0.00824
ре	411	0.00793
он	408	0.00787
ан	406	0.00784
ка	385	0.00743
од	366	0.00706
во	364	0.00703
те	352	0.00679
ал	352	0.00679
ав	344	0.00664
ас	342	0.00660
ат	339	0.00654
ор	338	0.00652
ем	338	0.00652
ес	338	0.00652

ес	338	0.00652
та	338	0.00652
ак	329	0.00635
ис	326	0.00629
ли	324	0.00625
ол	321	0.00620
ть	309	0.00596
да	307	0.00593
ле	302	0.00583
ло	300	0.00579
ог	300	0.00579
де	297	0.00573
ве	295	0.00569
ин	295	0.00569
за	294	0.00567
ся	293	0.00565
ит	290	0.00560
ла	282	0.00544
об	280	0.00540
ри	278	0.00537
мо	278	0.00537
же	271	0.00523
ел	264	0.00510
ме	263	0.00508
ед	259	0.00500
ив	257	0.00496
ой	247	0.00477
ти	247	0.00477
ль	239	0.00461
им	238	0.00459
ам	230	0.00444
ск	227	0.00438
ил	224	0.00432
оп	216	0.00417
ми	216	0.00417
сл	215	0.00415

сл	215	0.00415
до	215	0.00415
ки	213	0.00411
ег	213	0.00411
ны	206	0.00398
ви	201	0.00388
нн	200	0.00386
че	199	0.00384
тр	196	0.00378
ок	195	0.00376
аз	191	0.00369
ик	185	0.00357
ев	180	0.00347
еп	180	0.00347
ну	177	0.00342
се	176	0.00340
из	176	0.00340
со	175	0.00338
вс	175	0.00338
ад	174	0.00336
ру	173	0.00334
сп	173	0.00334
бы	172	0.00332
ож	171	0.00330
ае	171	0.00330
тс	169	0.00326
ар	165	0.00318
ди	165	0.00318
ез	164	0.00317
ьн	160	0.00309
ап	159	0.00307
ус	157	0.00303
мн	157	0.00303
чт	156	0.00301
тв	154	0.00297
сь	152	0.00293

сь	152	0.00293
ма	152	0.00293
ои	152	0.00293
дн	149	0.00288
ие	147	0.00284
вы	147	0.00284
ое	146	0.00282
ип	146	0.00282
кр	145	0.00280
уд	143	0.00276
ча	143	0.00276
бо	142	0.00274
си	141	0.00272
ид	140	0.00270
му	138	0.00266
ир	136	0.00262
ая	135	0.00261
ей	132	0.00255
аю	131	0.00253
хо	126	0.00243
яв	125	0.00241
ля	124	0.00239
ше	122	0.00235
ут	122	0.00235
оч	122	0.00235
пе	121	0.00234
ьв	121	0.00234
вн	119	0.00230
оз	117	0.00226
еб	116	0.00224
тн	114	0.00220
уж	114	0.00220
оо	113	0.00218
са	112	0.00216
ши	111	0.00214
их	110	0.00212

их	110	0.00212
эт	110	0.00212
ый	109	0.00210
ту	109	0.00210
ий	108	0.00208
бе	108	0.00208
ук	108	0.00208
йс	108	0.00208
ян	107	0.00207
ня	107	0.00207
ек	105	0.00203
еч	104	0.00201
ую	104	0.00201
чи	103	0.00199
яс	103	0.00199
па	101	0.00195
нт	100	0.00193
ич	100	0.00193
ьк	97	0.00187
ьс	96	0.00185
ео	96	0.00185
гл	96	0.00185
ее	96	0.00185
ят	94	0.00181
св	93	0.00179
лу	90	0.00174
ще	90	0.00174
ку	89	0.00172
ун	89	0.00172
эн	88	0.00170
рт	88	0.00170
вл	87	0.00168
жи	87	0.00168
бр	85	0.00164
ыл	85	0.00164
ым	85	0.00164

ым	85	0.00164	ые	64	0.00124
сн	85	0.00164	лс	64	0.00124
аж	84	0.00162	ты	64	0.00124
ду	84	0.00162	еш	64	0.00124
ия	83	0.00160	уч	64	0.00124
дв	83	0.00160	гд	64	0.00124
нг	81	0.00156	сс	64	0.00124
аб	81	0.00156	вп	63	0.00122
мп	80	0.00154	ув	63	0.00122
ву	80	0.00154	бу	63	0.00122
пл	79	0.00152	яд	63	0.00122
ао	79	0.00152	щи	62	0.00120
кв	79	0.00152	лю	61	0.00118
вш	78	0.00151	жа	61	0.00118
чн	74	0.00143	мы	61	0.00118
др	74	0.00143	вр	61	0.00118
ры	74	0.00143	би	60	0.00116
ио	73	0.00141	эв	59	0.00114
ач	72	0.00139	кт	58	0.00112
ах	72	0.00139	мс	58	0.00112
эд	71	0.00137	еж	57	0.00110
иб	70	0.00135	нь	56	0.00108
яе	70	0.00135	ке	56	0.00108
ум	70	0.00135	пи	56	0.00108
рн	70	0.00135	лн	55	0.00106
еи	69	0.00133	зо	55	0.00106
тк	69	0.00133	ци	55	0.00106
ии	69	0.00133	як	54	0.00104
уп	69	0.00133	ря	53	0.00102
оя	68	0.00131	ьв	53	0.00102
яп	68	0.00131	ых	53	0.00102
ют	67	0.00129	нд	53	0.00102
вк	66	0.00127	см	52	0.00100
га	65	0.00125	це	52	0.00100
аи	65	0.00125	нс	52	0.00100
ые	64	0.00124	еу	51	0.00098

еу	51	0.00098
аг	50	0.00097
вт	50	0.00097
ын	50	0.00097
ул	50	0.00097
ыс	50	0.00097
яо	49	0.00095
лы	49	0.00095
уб	49	0.00095
гв	48	0.00093
яз	48	0.00093
ош	48	0.00093
гр	48	0.00093
йв	47	0.00091
бл	47	0.00091
яи	47	0.00091
кл	47	0.00091
ьп	47	0.00091
тп	46	0.00089
иа	46	0.00089
аш	46	0.00089
мв	46	0.00089
йн	45	0.00087
ещ	45	0.00087
кн	45	0.00087
ур	44	0.00085
йд	44	0.00085
уг	44	0.00085
оу	44	0.00085
жд	44	0.00085
ша	44	0.00085
юс	44	0.00085
мк	43	0.00083
су	43	0.00083
йп	43	0.00083
ца	43	0.00083

ца	43	0.00083
иц	43	0.00083
жн	42	0.00081
ям	42	0.00081
ью	41	0.00079
ох	40	0.00077
пу	39	0.00075
ыт	39	0.00075
ып	39	0.00075
ье	39	0.00075
йк	39	0.00075
ба	38	0.00073
тя	37	0.00071
гу	37	0.00071
иг	37	0.00071
мя	37	0.00071
еа	37	0.00071
дь	37	0.00071
иу	36	0.00069
ющ	36	0.00069
уш	36	0.00069
аа	36	0.00069
сч	36	0.00069
ьи	36	0.00069
дс	35	0.00068
ща	35	0.00068
ай	35	0.00068
шн	34	0.00066
йч	34	0.00066
йт	34	0.00066
сё	34	0.00066
юя	34	0.00066
рь	33	0.00064
кс	33	0.00064
яб	33	0.00064
вя	33	0.00064

вя	33	0.00064
зу	33	0.00064
ял	33	0.00064
яу	33	0.00064
зе	32	0.00062
лк	32	0.00062
уо	32	0.00062
ьд	32	0.00062
ды	32	0.00062
йо	32	0.00062
вз	32	0.00062
ец	31	0.00060
ау	31	0.00060
рк	31	0.00060
уе	31	0.00060
лв	31	0.00060
зы	31	0.00060
рв	30	0.00058
ср	30	0.00058
хв	30	0.00058
хн	30	0.00058
жу	30	0.00058
эг	30	0.00058
ех	30	0.00058
хи	29	0.00056
ги	29	0.00056
дя	29	0.00056
сд	29	0.00056
ьо	29	0.00056
ьз	29	0.00056
вд	29	0.00056
тт	29	0.00056
дл	29	0.00056
ию	28	0.00054
ьш	28	0.00054
бу	28	0.00054

ьу	28	0.00054
тм	28	0.00054
эм	28	0.00054
ящ	27	0.00052
сы	27	0.00052
юд	27	0.00052
юп	27	0.00052
иж	27	0.00052
мм	27	0.00052
йи	27	0.00052
иш	27	0.00052
уи	27	0.00052
оэ	27	0.00052
зи	27	0.00052
йр	27	0.00052
вм	26	0.00050
юн	26	0.00050
яч	26	0.00050
нк	26	0.00050
аз	25	0.00048
ьч	25	0.00048
яр	25	0.00048
ух	25	0.00048
мт	25	0.00048
ьч	25	0.00048
нц	24	0.00046
хп	24	0.00046
ьт	24	0.00046
шь	24	0.00046
эл	24	0.00046
бн	24	0.00046
хс	24	0.00046
чу	23	0.00044
йм	23	0.00044
лч	23	0.00044
ьм	23	0.00044

ьм	23	0.00044
юч	23	0.00044
ьр	23	0.00044
бщ	23	0.00044
ац	22	0.00042
мд	22	0.00042
дк	22	0.00042
ащ	22	0.00042
ея	22	0.00042
фо	21	0.00041
яг	21	0.00041
цы	21	0.00041
ьо	21	0.00041
эр	21	0.00041
тб	21	0.00041
ьб	21	0.00041
ык	21	0.00041
ха	21	0.00041
чк	21	0.00041
рх	20	0.00039
мл	20	0.00039
шл	20	0.00039
оа	20	0.00039
вв	20	0.00039
шк	20	0.00039
ьш	20	0.00039
эк	20	0.00039
тч	20	0.00039
рю	20	0.00039
йз	19	0.00037
ьи	19	0.00037
юб	19	0.00037
лп	19	0.00037
мб	19	0.00037
мр	19	0.00037
ьб	19	0.00037

-		
ьб	19	0.00037
тд	19	0.00037
яю	19	0.00037
мч	18	0.00035
ыр	18	0.00035
хр	18	0.00035
шу	18	0.00035
рш	17	0.00033
хк	17	0.00033
цо	17	0.00033
ья	17	0.00033
иф	17	0.00033
ье	17	0.00033
эп	16	0.00031
тл	16	0.00031
пы	16	0.00031
уа	16	0.00031
ьу	16	0.00031
юв	16	0.00031
шо	16	0.00031
йб	16	0.00031
бя	16	0.00031
гн	16	0.00031
сх	16	0.00031
уз	16	0.00031
кц	16	0.00031
фи	16	0.00031
пъ	16	0.00031
йц	16	0.00031
рм	15	0.00029
ьд	15	0.00029
лж	15	0.00029
кп	15	0.00029
чь	15	0.00029
нв	15	0.00029
кк	15	0.00029

кк	15	0.00029
мг	15	0.00029
дп	15	0.00029
рс	15	0.00029
рп	15	0.00029
пп	15	0.00029
еф	15	0.00029
йг	14	0.00027
ьа	14	0.00027
юе	14	0.00027
юк	14	0.00027
яа	14	0.00027
вг	14	0.00027
хл	14	0.00027
юи	14	0.00027
рг	14	0.00027
нп	13	0.00025
ыг	13	0.00025
хд	13	0.00025
лд	13	0.00025
ущ	13	0.00025
уя	13	0.00025
ге	13	0.00025
аф	13	0.00025
юм	13	0.00025
юо	12	0.00023
эс	12	0.00023
дм	12	0.00023
лл	12	0.00023
оц	12	0.00023
нч	12	0.00023
уй	12	0.00023
кд	12	0.00023
цу	12	0.00023
оф	12	0.00023
уу	12	0.00023

уу	12	0.00023
эк	12	0.00023
вч	12	0.00023
ощ	12	0.00023
хм	11	0.00021
дц	11	0.00021
рж	11	0.00021
пя	11	0.00021
йл	11	0.00021
из	11	0.00021
яя	11	0.00021
йу	11	0.00021
ху	11	0.00021
ьг	11	0.00021
юз	11	0.00021
вэ	11	0.00021
юр	11	0.00021
сш	11	0.00021
вь	10	0.00019
яж	10	0.00019
ищ	10	0.00019
пн	10	0.00019
нз	10	0.00019
ьл	10	0.00019
ыз	10	0.00019
лг	9	0.00017
рд	9	0.00017
зб	9	0.00017
кз	9	0.00017
ьц	9	0.00017
хт	9	0.00017
йж	9	0.00017
рч	9	0.00017
щу	9	0.00017
йш	9	0.00017
фу	9	0.00017

фУ	9	0.00017
дТ	9	0.00017
еЮ	9	0.00017
мЗ	9	0.00017
сЗ	9	0.00017
бС	9	0.00017
вБ	9	0.00017
кЖ	9	0.00017
дЪ	9	0.00017
тГ	9	0.00017
яЦ	8	0.00015
лБ	8	0.00015
йЕ	8	0.00015
кБ	8	0.00015
пК	8	0.00015
цВ	8	0.00015
оЮ	8	0.00015
ёВ	8	0.00015
рЦ	8	0.00015
эЯ	8	0.00015
сБ	8	0.00015
йА	8	0.00015
эЗ	8	0.00015
еЭ	8	0.00015
сЮ	7	0.00014
эЖ	7	0.00014
лТ	7	0.00014
юГ	7	0.00014
юЛ	7	0.00014
вХ	7	0.00014
юА	7	0.00014
кМ	7	0.00014
мХ	7	0.00014
фЕ	7	0.00014
ыЯ	7	0.00014
хЗ	7	0.00014

хэ	7	0.00014
дш	7	0.00014
дб	7	0.00014
вж	7	0.00014
нр	6	0.00012
ьх	6	0.00012
ыа	6	0.00012
мж	6	0.00012
лэ	6	0.00012
чо	6	0.00012
фа	6	0.00012
фт	6	0.00012
сг	6	0.00012
яэ	6	0.00012
кч	6	0.00012
тш	6	0.00012
ях	5	0.00010
лр	5	0.00010
тж	5	0.00010
гк	5	0.00010
фр	5	0.00010
йэ	5	0.00010
рл	5	0.00010
ёэ	5	0.00010
йя	5	0.00010
нж	5	0.00010
сэ	5	0.00010
дж	5	0.00010
жб	5	0.00010
тэ	5	0.00010
мэ	5	0.00010
бэ	5	0.00010
нб	5	0.00010
иы	5	0.00010
лм	5	0.00010
нф	5	0.00010

нф	5	0.00010
кэ	5	0.00010
нщ	5	0.00010
яш	5	0.00010
сц	5	0.00010
эт	5	0.00010
дд	5	0.00010
тц	5	0.00010
гс	4	0.00008
ню	4	0.00008
хе	4	0.00008
яы	4	0.00008
тх	4	0.00008
гп	4	0.00008
кш	4	0.00008
рр	4	0.00008
дэ	4	0.00008
кг	4	0.00008
тю	4	0.00008
юш	4	0.00008
уы	4	0.00008
ьж	4	0.00008
жч	4	0.00008
йф	4	0.00008
юу	4	0.00008
хб	4	0.00008
йх	4	0.00008
дч	4	0.00008
ьдж	4	0.00008
цп	4	0.00008
цк	3	0.00006
шв	3	0.00006
рэ	3	0.00006
эл	3	0.00006
хш	3	0.00006
дх	3	0.00006

ДХ	3	0.00006
БК	3	0.00006
ШТ	3	0.00006
ХЯ	3	0.00006
КЯ	3	0.00006
МЬ	3	0.00006
АЫ	3	0.00006
ЁН	3	0.00006
УЭ	3	0.00006
ЖМ	3	0.00006
НЛ	3	0.00006
МЦ	3	0.00006
ВЦ	3	0.00006
ФП	3	0.00006
НМ	3	0.00006
ТЭ	3	0.00006
ПС	3	0.00006
ЖС	3	0.00006
ПЦ	3	0.00006
ЦС	3	0.00006
БЪ	3	0.00006
ТЩ	3	0.00006
ЭН	3	0.00006
ЧШ	3	0.00006
ЧМ	3	0.00006
СЖ	2	0.00004
Ёе	2	0.00004
Эш	2	0.00004
ДГ	2	0.00004
Ёп	2	0.00004
Шм	2	0.00004
ГЭ	2	0.00004
ЖК	2	0.00004
Бю	2	0.00004
Бх	2	0.00004
Эф	2	0.00004

эф	2	0.00004
ёб	2	0.00004
жо	2	0.00004
юж	2	0.00004
бз	2	0.00004
эц	2	0.00004
хч	2	0.00004
уф	2	0.00004
мф	2	0.00004
ьы	2	0.00004
ёр	2	0.00004
ээ	2	0.00004
тъ	2	0.00004
ёл	2	0.00004
чл	2	0.00004
ёт	2	0.00004
лэ	2	0.00004
цч	2	0.00004
зч	2	0.00004
пд	2	0.00004
зъ	2	0.00004
гт	2	0.00004
цн	2	0.00004
щд	2	0.00004
фы	2	0.00004
пч	2	0.00004
хф	2	0.00004
цт	2	0.00004
хы	2	0.00004
хж	2	0.00004
ьц	2	0.00004
шп	2	0.00004
лх	2	0.00004
гб	2	0.00004
ьэ	2	0.00004
бц	2	0.00004

бц	2	0.00004
ьф	1	0.00002
бш	1	0.00002
цш	1	0.00002
бт	1	0.00002
фв	1	0.00002
сщ	1	0.00002
мш	1	0.00002
ёч	1	0.00002
ью	1	0.00002
гю	1	0.00002
чв	1	0.00002
жв	1	0.00002
вщ	1	0.00002
фм	1	0.00002
бь	1	0.00002
юх	1	0.00002
фк	1	0.00002
съ	1	0.00002
юэ	1	0.00002
яф	1	0.00002
фч	1	0.00002
фн	1	0.00002
йю	1	0.00002
сф	1	0.00002
гж	1	0.00002
ьъ	1	0.00002
ёу	1	0.00002
гм	1	0.00002
эс	1	0.00002
фф	1	0.00002
эь	1	0.00002
эр	1	0.00002
уц	1	0.00002
бм	1	0.00002
лф	1	0.00002

лф	1	0.00002
тф	1	0.00002
пю	1	0.00002
ёж	1	0.00002
ём	1	0.00002
пб	1	0.00002
ёд	1	0.00002
юю	1	0.00002
щъ	1	0.00002
шр	1	0.00002
жл	1	0.00002
ыф	1	0.00002
шг	1	0.00002
ьы	1	0.00002
юф	1	0.00002
дэ	1	0.00002
хэ	1	0.00002
мщ	1	0.00002
еы	1	0.00002
ёк	1	0.00002
зэ	1	0.00002
гш	1	0.00002
фл	1	0.00002
нш	1	0.00002
шэ	1	0.00002
цр	1	0.00002
чп	1	0.00002
пт	1	0.00002
бд	1	0.00002
бп	1	0.00002
йй	1	0.00002
жц	1	0.00002
жь	1	0.00002
ъя	1	0.00002
цм	1	0.00002
шя	1	0.00002
шя	1	0.00002
гч	1	0.00002

Частота біграм в тексті без перетинів:

Символ	Кількість	Частота
то	374	0.01444
на	324	0.01251
но	309	0.01193
ст	293	0.01131
по	282	0.01089
ен	277	0.01069
ов	273	0.01054
ос	272	0.01050
ни	271	0.01046
ра	260	0.01004
ко	258	0.00996
ро	247	0.00953
не	234	0.00903
ом	226	0.00872
ет	224	0.00865
го	221	0.00853
ер	216	0.00834
ва	214	0.00826
пр	211	0.00814
ан	209	0.00807
от	209	0.00807
ре	208	0.00803
ка	195	0.00753
од	194	0.00749
те	192	0.00741
он	190	0.00733
ал	182	0.00703
во	180	0.00695
ть	172	0.00664
ор	171	0.00660
ав	170	0.00656
та	170	0.00656
ак	169	0.00652
ли	167	0.00645
ас	165	0.00637
ем	164	0.00633
ис	163	0.00629
ат	162	0.00625
ле	162	0.00625
ог	160	0.00618

ог	160	0.00618
ло	157	0.00606
ри	153	0.00591
ол	148	0.00571
да	147	0.00567
ве	144	0.00556
ес	143	0.00552
об	142	0.00548
де	142	0.00548
за	142	0.00548
ся	141	0.00544
ин	138	0.00533
ив	138	0.00533
мо	135	0.00521
ед	133	0.00513
ла	133	0.00513
ме	130	0.00502
ит	130	0.00502
ль	129	0.00498
же	125	0.00482
ой	124	0.00479
ел	120	0.00463
ам	118	0.00455
оп	117	0.00452
ти	116	0.00448
ки	111	0.00428
че	109	0.00421
им	109	0.00421
ми	109	0.00421
сл	107	0.00413
до	106	0.00409
ви	103	0.00398
ег	103	0.00398
ил	103	0.00398
ск	102	0.00394
ок	100	0.00386
нн	100	0.00386
ны	99	0.00382
тр	94	0.00363
ев	90	0.00347
ож	90	0.00347

ож	90	0.00347
ну	88	0.00340
еп	88	0.00340
ик	88	0.00340
ие	87	0.00336
из	87	0.00336
аз	87	0.00336
вс	87	0.00336
тс	87	0.00336
со	86	0.00332
ар	85	0.00328
сп	85	0.00328
ае	85	0.00328
ап	84	0.00324
ад	84	0.00324
ез	84	0.00324
се	81	0.00313
бы	81	0.00313
вы	80	0.00309
ид	79	0.00305
ир	78	0.00301
мн	78	0.00301
ип	78	0.00301
ди	77	0.00297
ру	76	0.00293
ая	75	0.00289
ма	75	0.00289
ьн	75	0.00289
ус	74	0.00286
дн	74	0.00286
сь	72	0.00278
ей	72	0.00278
бо	71	0.00274
ое	69	0.00266
уд	69	0.00266
оч	67	0.00259
ян	67	0.00259
ои	67	0.00259
си	67	0.00259
тв	66	0.00255
кр	66	0.00255

кр	66	0.00255
ча	66	0.00255
яв	64	0.00247
тн	63	0.00243
эт	63	0.00243
му	63	0.00243
еб	63	0.00243
чт	62	0.00239
ту	62	0.00239
ут	61	0.00235
оо	61	0.00235
аю	61	0.00235
пе	60	0.00232
бе	60	0.00232
уж	60	0.00232
йс	60	0.00232
еч	59	0.00228
ля	59	0.00228
па	58	0.00224
ук	58	0.00224
са	55	0.00212
хо	55	0.00212
нт	54	0.00208
ят	54	0.00208
ьв	54	0.00208
ий	53	0.00205
ек	53	0.00205
ее	53	0.00205
сн	53	0.00205
их	53	0.00205
оз	52	0.00201
ши	52	0.00201
вн	51	0.00197
ую	51	0.00197
ня	51	0.00197
чи	51	0.00197
зн	50	0.00193
яс	50	0.00193
ьк	50	0.00193
ду	50	0.00193
лу	49	0.00189

лу	49	0.00189
гл	49	0.00189
св	48	0.00185
ым	48	0.00185
ео	47	0.00181
ше	47	0.00181
ще	47	0.00181
аб	47	0.00181
ьс	46	0.00178
ый	46	0.00178
аж	46	0.00178
ыл	46	0.00178
жи	45	0.00174
нг	44	0.00170
чн	43	0.00166
ку	43	0.00166
вл	43	0.00166
ах	43	0.00166
ао	43	0.00166
рт	43	0.00166
ио	42	0.00162
сс	42	0.00162
ии	42	0.00162
бр	42	0.00162
тк	41	0.00158
вш	40	0.00154
дв	40	0.00154
ун	40	0.00154
кв	39	0.00151
ву	39	0.00151
ия	39	0.00151
мп	39	0.00151
эд	38	0.00147
ич	38	0.00147
ры	37	0.00143
уп	37	0.00143
ют	37	0.00143
лю	36	0.00139
еш	36	0.00139
ум	36	0.00139
бу	36	0.00139

бу	36	0.00139
пл	36	0.00139
яе	36	0.00139
оя	35	0.00135
га	35	0.00135
рн	35	0.00135
ты	34	0.00131
лс	34	0.00131
яд	34	0.00131
эв	34	0.00131
уч	34	0.00131
ые	33	0.00127
еи	33	0.00127
вр	33	0.00127
др	33	0.00127
жа	33	0.00127
см	32	0.00124
ач	32	0.00124
це	31	0.00120
гд	31	0.00120
аи	31	0.00120
нь	30	0.00116
лы	30	0.00116
кт	30	0.00116
ьс	30	0.00116
ош	30	0.00116
еу	30	0.00116
мв	30	0.00116
ув	29	0.00112
вп	29	0.00112
пи	29	0.00112
вк	29	0.00112
щи	29	0.00112
яп	28	0.00108
лн	28	0.00108
мы	28	0.00108
яи	28	0.00108
иб	28	0.00108
би	28	0.00108
еж	27	0.00104
аг	27	0.00104

аг	27	0.00104
ьп	27	0.00104
нд	27	0.00104
йд	27	0.00104
мс	27	0.00104
пу	26	0.00100
йк	26	0.00100
гр	26	0.00100
ьв	25	0.00096
тп	25	0.00096
ке	25	0.00096
су	25	0.00096
юс	25	0.00096
иа	25	0.00096
зо	25	0.00096
як	25	0.00096
йв	24	0.00093
ын	24	0.00093
уб	24	0.00093
аш	24	0.00093
гу	23	0.00089
бл	23	0.00089
яо	23	0.00089
ых	23	0.00089
йн	23	0.00089
ул	23	0.00089
кл	23	0.00089
йп	23	0.00089
ша	23	0.00089
ох	23	0.00089
ца	23	0.00089
нс	23	0.00089
ря	22	0.00085
вт	22	0.00085
ещ	22	0.00085
ям	22	0.00085
кн	22	0.00085
ющ	21	0.00081
ье	21	0.00081
уг	21	0.00081
жд	21	0.00081

жд	21	0.00081
кс	20	0.00077
ур	20	0.00077
гв	20	0.00077
сч	20	0.00077
юя	20	0.00077
ып	20	0.00077
иц	20	0.00077
дь	20	0.00077
зы	20	0.00077
зе	19	0.00073
иг	19	0.00073
ыт	19	0.00073
жн	19	0.00073
йт	19	0.00073
рк	19	0.00073
дл	19	0.00073
яб	18	0.00069
шн	18	0.00069
уш	18	0.00069
яз	18	0.00069
ды	18	0.00069
жу	18	0.00069
ци	18	0.00069
зг	18	0.00069
мк	18	0.00069
рь	18	0.00069
ех	18	0.00069
оу	18	0.00069
ец	17	0.00066
рв	17	0.00066
ги	17	0.00066
вя	17	0.00066
ию	17	0.00066
тт	17	0.00066
аа	17	0.00066
ью	17	0.00066
ьи	17	0.00066
еа	17	0.00066
йо	17	0.00066
вз	17	0.00066

вэ	17	0.00066
сы	17	0.00066
ба	17	0.00066
хс	17	0.00066
ьч	17	0.00066
иу	17	0.00066
хи	16	0.00062
лк	16	0.00062
эу	16	0.00062
хн	16	0.00062
мм	16	0.00062
мя	16	0.00062
ща	16	0.00062
ср	15	0.00058
юд	15	0.00058
ял	15	0.00058
цы	15	0.00058
тя	15	0.00058
яу	15	0.00058
йр	15	0.00058
дя	15	0.00058
сё	15	0.00058
чк	15	0.00058
ьо	14	0.00054
нц	14	0.00054
ац	14	0.00054
йи	14	0.00054
иш	14	0.00054
шь	14	0.00054
ьу	14	0.00054
зи	14	0.00054
тм	14	0.00054
вд	14	0.00054
мд	14	0.00054
дс	14	0.00054
ящ	13	0.00050
мб	13	0.00050
ьо	13	0.00050
эл	13	0.00050
мт	13	0.00050
уе	13	0.00050

уе	13	0.00050
эм	13	0.00050
нк	13	0.00050
мл	12	0.00046
иж	12	0.00046
йч	12	0.00046
хп	12	0.00046
хв	12	0.00046
ух	12	0.00046
лв	12	0.00046
шо	12	0.00046
юн	12	0.00046
ьч	12	0.00046
ьэ	12	0.00046
ащ	12	0.00046
уз	12	0.00046
сд	12	0.00046
бщ	12	0.00046
кц	12	0.00046
ау	12	0.00046
аз	12	0.00046
мч	11	0.00042
мр	11	0.00042
вв	11	0.00042
ьб	11	0.00042
юч	11	0.00042
ьр	11	0.00042
яг	11	0.00042
ай	11	0.00042
шу	11	0.00042
иф	11	0.00042
йэ	11	0.00042
яр	10	0.00039
чу	10	0.00039
рх	10	0.00039
ьм	10	0.00039
тч	10	0.00039
тд	10	0.00039
бн	10	0.00039
яч	10	0.00039
ьд	10	0.00039
аа	9	0.00035

ьд	10	0.00039
шл	10	0.00039
ьш	10	0.00039
уи	10	0.00039
еф	10	0.00039
вм	10	0.00039
вч	10	0.00039
ея	10	0.00039
уо	10	0.00039
лч	10	0.00039
кк	9	0.00035
хк	9	0.00035
лп	9	0.00035
ыр	9	0.00035
бя	9	0.00035
ья	9	0.00035
эр	9	0.00035
ык	9	0.00035
юв	9	0.00035
рю	9	0.00035
юп	9	0.00035
эк	9	0.00035
йм	9	0.00035
дк	9	0.00035
рш	9	0.00035
пы	8	0.00031
нп	8	0.00031
уа	8	0.00031
нв	8	0.00031
ыг	8	0.00031
оэ	8	0.00031
иэ	8	0.00031
чь	8	0.00031
юб	8	0.00031
тл	8	0.00031
юе	8	0.00031
хл	8	0.00031
юи	8	0.00031
цу	8	0.00031
сш	8	0.00031
оф	8	0.00031

оф	8	0.00031
эс	8	0.00031
оа	8	0.00031
сх	8	0.00031
пь	8	0.00031
фи	8	0.00031
фо	8	0.00031
эк	8	0.00031
ье	8	0.00031
йг	7	0.00027
хм	7	0.00027
ьи	7	0.00027
хд	7	0.00027
лд	7	0.00027
яа	7	0.00027
дп	7	0.00027
вх	7	0.00027
ьт	7	0.00027
яю	7	0.00027
мг	7	0.00027
уя	7	0.00027
кп	7	0.00027
юм	7	0.00027
нч	7	0.00027
юр	7	0.00027
рг	7	0.00027
шк	7	0.00027
йц	7	0.00027
ьб	7	0.00027
ьу	7	0.00027
хр	7	0.00027
эп	6	0.00023
эб	6	0.00023
хт	6	0.00023
оц	6	0.00023
йб	6	0.00023
рж	6	0.00023
лж	6	0.00023
щу	6	0.00023
ою	6	0.00023
яя	6	0.00023

яя	6	0.00023
рм	6	0.00023
пн	6	0.00023
кд	6	0.00023
зя	6	0.00023
йу	6	0.00023
пя	6	0.00023
уу	6	0.00023
сз	6	0.00023
бс	6	0.00023
дъ	6	0.00023
зз	6	0.00023
дм	6	0.00023
ге	6	0.00023
ущ	6	0.00023
ьш	6	0.00023
нз	6	0.00023
яц	5	0.00019
кз	5	0.00019
юо	5	0.00019
вь	5	0.00019
йл	5	0.00019
ищ	5	0.00019
лб	5	0.00019
юк	5	0.00019
пк	5	0.00019
дц	5	0.00019
йш	5	0.00019
тб	5	0.00019
фу	5	0.00019
сб	5	0.00019
яж	5	0.00019
мз	5	0.00019
аф	5	0.00019
пп	5	0.00019
ёв	5	0.00019
йа	5	0.00019
ба	5	0.00019
ху	5	0.00019
вэ	5	0.00019
ьц	5	0.00019

ьц	5	0.00019
нр	4	0.00015
ьд	4	0.00015
эж	4	0.00015
ьх	4	0.00015
лг	4	0.00015
сю	4	0.00015
гн	4	0.00015
вг	4	0.00015
мж	4	0.00015
дт	4	0.00015
ьг	4	0.00015
рп	4	0.00015
дж	4	0.00015
тэ	4	0.00015
вб	4	0.00015
рс	4	0.00015
ею	4	0.00015
тш	4	0.00015
яэ	4	0.00015
ощ	4	0.00015
эт	4	0.00015
чо	4	0.00015
рч	4	0.00015
рд	3	0.00012
хе	3	0.00012
лт	3	0.00012
гк	3	0.00012
лл	3	0.00012
шв	3	0.00012
фр	3	0.00012
хш	3	0.00012
йэ	3	0.00012
юл	3	0.00012
бк	3	0.00012
лэ	3	0.00012
мх	3	0.00012
уй	3	0.00012
фа	3	0.00012
ха	3	0.00012
ьа	3	0.00012

ыа	3	0.00012
ыз	3	0.00012
кж	3	0.00012
мь	3	0.00012
юг	3	0.00012
тю	3	0.00012
нб	3	0.00012
юа	3	0.00012
сг	3	0.00012
цв	3	0.00012
лм	3	0.00012
нл	3	0.00012
дш	3	0.00012
йя	3	0.00012
км	3	0.00012
юу	3	0.00012
йх	3	0.00012
сц	3	0.00012
вж	3	0.00012
юз	3	0.00012
тщ	3	0.00012
жч	3	0.00012
ьж	3	0.00012
цо	3	0.00012
цп	3	0.00012
ее	2	0.00008
эш	2	0.00008
дг	2	0.00008
еп	2	0.00008
яы	2	0.00008
рц	2	0.00008
рл	2	0.00008
гп	2	0.00008
хя	2	0.00008
кя	2	0.00008
кш	2	0.00008
жк	2	0.00008
рр	2	0.00008
фе	2	0.00008
жо	2	0.00008
ыя	2	0.00008

ыя	2	0.00008
уэ	2	0.00008
ьл	2	0.00008
лр	2	0.00008
жм	2	0.00008
тг	2	0.00008
мэ	2	0.00008
пс	2	0.00008
ьж	2	0.00008
нщ	2	0.00008
кэ	2	0.00008
яш	2	0.00008
йе	2	0.00008
ёл	2	0.00008
чл	2	0.00008
аы	2	0.00008
цс	2	0.00008
хб	2	0.00008
фт	2	0.00008
лэ	2	0.00008
нф	2	0.00008
йж	2	0.00008
дд	2	0.00008
пд	2	0.00008
кч	2	0.00008
дб	2	0.00008
цк	2	0.00008
жб	2	0.00008
чш	2	0.00008
бэ	2	0.00008
ях	2	0.00008
нм	2	0.00008
кб	2	0.00008
нж	2	0.00008
тц	2	0.00008
ьф	1	0.00004
тх	1	0.00004
фв	1	0.00004
сщ	1	0.00004
мш	1	0.00004
ёч	1	0.00004
шч	1	0.00004

ёч	1	0.00004
шм	1	0.00004
чв	1	0.00004
рз	1	0.00004
бю	1	0.00004
бх	1	0.00004
эф	1	0.00004
ёб	1	0.00004
юш	1	0.00004
дх	1	0.00004
юх	1	0.00004
фк	1	0.00004
сѣ	1	0.00004
юэ	1	0.00004
яф	1	0.00004
фч	1	0.00004
фн	1	0.00004
сф	1	0.00004
гж	1	0.00004
ёу	1	0.00004
жс	1	0.00004
эс	1	0.00004
пц	1	0.00004
йф	1	0.00004
фф	1	0.00004
эь	1	0.00004
хз	1	0.00004
гс	1	0.00004
ээ	1	0.00004
эр	1	0.00004
тѣ	1	0.00004
бм	1	0.00004
пю	1	0.00004
уы	1	0.00004
цч	1	0.00004
ёж	1	0.00004
сэ	1	0.00004
гт	1	0.00004
бѣ	1	0.00004
ню	1	0.00004
иы	1	0.00004

иы	1	0.00004
шр	1	0.00004
жл	1	0.00004
еэ	1	0.00004
ьы	1	0.00004
ёт	1	0.00004
ён	1	0.00004
эн	1	0.00004
мф	1	0.00004
ьы	1	0.00004
пч	1	0.00004
дэ	1	0.00004
хж	1	0.00004
ыц	1	0.00004
уф	1	0.00004
хф	1	0.00004
еы	1	0.00004
ёк	1	0.00004
эч	1	0.00004
тэ	1	0.00004
зэ	1	0.00004
тж	1	0.00004
мц	1	0.00004
шэ	1	0.00004
шт	1	0.00004
чп	1	0.00004
пт	1	0.00004
бд	1	0.00004
бп	1	0.00004
гб	1	0.00004
бц	1	0.00004
жц	1	0.00004
ъя	1	0.00004
цм	1	0.00004
ёр	1	0.00004
дч	1	0.00004
хч	1	0.00004
фы	1	0.00004
шя	1	0.00004
шп	1	0.00004

Для обрахунку $H1$ використовую формулу: $H1 = -\sum_{i=1}^n q_i \log_2 q_i$, де q – частота, n – кількість символів. Для $H2$: $H2 = -\frac{1}{2} \sum_{i,j} q_{ij} \log_2 q_{ij}$.

Отримані значення ентропії:

Ентропія для символів з пробілами: 4.472826862297906
 Ентропія для символів без пробілів: 4.472826862297906
 Ентропія для біграм з пробілами і перетинами: 8.320299477822019
 Ентропія для біграм з пробілами без перетинів: 8.312355987491514
 Ентропія для біграм без пробілів з перетинами: 8.320299477822019
 Ентропія для біграм без пробілів без перетинів: 8.312355987491514

Надлишковість рахується за формулою $R = 1 - \frac{H_\infty}{H_0}$; H_∞ дорівнюватиме значенню, порахованому вище, а $H_0 = \log_2 m$, де m – кількість символів у алфавіті (для алфавіту з пробілами: $m = 34$; без: $m = 33$).

Отримані значення надлишковості:

Надлишковість для символів з пробілами: 0.11330741483245543
 Надлишковість для символів без пробілів: 0.257297627584014
 Надлишковість для біграм з пробілами і перетинами: 0.17529248498923833
 Надлишковість для біграм з пробілами без перетинів: 0.17607984320734626
 Надлишковість для біграм без пробілів з перетинами: 0.3092169279479786
 Надлишковість для біграм без пробілів без перетинів: 0.30987642688403827

	Ентропія	Надлишковість
Букви з пробілами	4,472827	0,113307
Букви без пробілів	4,472826	0,257297
Біграми з пробілами та перетинами	8,320299	0,175292
Біграми з пробілами та без перетинів	8,312356	0,176080
Біграми без пробілів та з перетинами	8,320299	0,309217
Біграми без пробілів та перетинів	8,312356	0,309876

CoolPinkProgram:

H(10):

