



***Міністерство освіти і науки України
НТУУ «Київський політехнічний інститут ім. Ігоря
Сікорського»
Навчально-науковий Фізико-технічний інститут***

Криптографія
Комп'ютерний практикум №1.
Експериментальна оцінка ентропії на символ
джерела відкритого тексту
Варіант №3

ВИКОНАВ:

СТУДЕНТ ГРУПИ ФБ-31

ВАСАЛАТІЙ А. Ю., ЯКОВЧУК О. С.

ПЕРЕВІРЕНО:

БАЙДЕНКО П. В.

17.09.2025

Початково було імплементовано обрахунок частот монограм та біграм у тексті за формулою:

$v(a) = \frac{N_a}{N}$, де $v(a)$ — власне частота n -грами a ; N_a - кількість її появ у тексті, а N - кількість n -грам у тексті загалом.

Для біграм було враховано два випадки - як з можливістю перетину букв, так і без неї. Крім того при обрахунках завжди розглядалися два варіанти в залежності від наявності пробілу в використовуваному для досліджень алфавіті.

Текст перед виконанням обчислень був приведений у нижній регістр, а літери ё та ъ були замінені на е та ь відповідно (за аналогією до наданої нам CoolPinkProgram.exe). Символи чи їх послідовності, що не належать до алфавіту видалялися або замінювалися на 1 пробіл в залежності від його наявності в алфавіті, що використовувався

(з технічних моментів варто відзначити, що було використано модулі **python-magic** <https://pypi.org/project/python-magic/> для перевірки переданого файлу на підходящий тип, а також **chardet** <https://pypi.org/project/chardet/>, щоб автоматично детектувати кодування файлу і уникнути необхідності редагувати код щоразу, коли необхідно опрацювати файл з кодування відмінним від попереднього (зіткнулися з подібною ситуацією, коли здійснювали тестування як на TEXT від CoolPinkProgram.exe з кодуванням IBM866, так і на нашому власному з UTF-8)).

Після цього для кожного переліку частот n -грам було обраховано ентропію, послуговуючись загальною формулою:

$$H_n = \frac{1}{n} H(x_1, x_2, \dots, x_n). \text{ Як приклад } H_1 = - \sum_{i=1}^n p_i \log_2 p_i.$$

Ймовірності, що присутні у формулах були замінені відповідними частотами.

Далі, базуючись на отриманих ентропіях, було обчислено надлишковість за формулою:

$R = H/H_0$, де $H_0 = \log_2 m$ (m - потужність алфавіту, що використовувався).

Результат роботи скрипта для файлу від CoolPinkProgram та “Анна Кареніна” Льва Толстого (3.2МБ):

```

(venv) the_old_man@theOldMan-machine:~/Desktop/crypto25-26/lab1/vasalatii_fb-31_yakovchuk_fb-31_cp1$ /home/the_old_man/Desktop/crypto25-26/venv/bin/python ./fb-31_yakovchuk_fb-31_cp1/lab1.py -f ../../tasks/cp1/TEXT FILE

TextAnalyzer

=====
File type of ../../tasks/cp1/TEXT: text/plain
Encoding detection result: {'encoding': 'IBM866', 'confidence': 0.99, 'language': 'Russian'}
Text length: 5759
Monograms and bigrams stats were saved to /home/the_old_man/Desktop/crypto25-26/lab1/vasalatii_fb-31_yakovchuk_fb-31_cp1/TEXT_stats.xlsx
H1: 4.422101851111231
H1 with ws: 4.385354297866487
H2 not overlapped: 3.9748832357629778
H2 not overlapped with ws: 3.8839674721838
H2 overlapping: 4.02077298377601
H2 overlapping with ws: 3.914743461832841
Redundancy based on monogram entropy: 0.10740278058018504
Redundancy based on monogram (including whitespaces) entropy: 0.11482024063676377
Redundancy based on bigram (not overlapped) entropy: 0.1076734495907414
Redundancy based on bigram (not overlapped, including whitespaces) entropy: 0.21602471342511265
Redundancy based on bigram (overlapping) entropy: 0.18641064587082012
Redundancy based on bigram (overlapping, including whitespaces) entropy: 0.2098126080257935
=====
File type of FILE: text/plain
Encoding detection result: {'encoding': 'utf-8', 'confidence': 0.99, 'language': ''}
Text length: 1778358
Monograms and bigrams stats were saved to /home/the_old_man/Desktop/crypto25-26/lab1/vasalatii_fb-31_yakovchuk_fb-31_cp1/FILE_stats.xlsx
H1: 4.4393304946163665
H1 with ws: 4.347574526370398
H2 not overlapped: 4.122434891829624
H2 not overlapped with ws: 3.9412994623955098
H2 overlapping: 4.121835034126119
H2 overlapping with ws: 3.9413951409046173
Redundancy based on monogram entropy: 0.10392519462562477
Redundancy based on monogram (including whitespaces) entropy: 0.12244605300452982
Redundancy based on bigram (not overlapped) entropy: 0.16789028259001282
Redundancy based on bigram (not overlapped, including whitespaces) entropy: 0.20445230356894506
Redundancy based on bigram (overlapping) entropy: 0.16801136331954447
Redundancy based on bigram (overlapping, including whitespaces) entropy: 0.20443299094927625
(venv) the_old_man@theOldMan-machine:~/Desktop/crypto25-26/lab1/vasalatii_fb-31_yakovchuk_fb-31_cp1$

```

Статистика по кількості входжень n-грам та їх частотам зберігається в окремий файл для кожного з вхідних файлів. Для цього використовуються **pandas** - <https://pypi.org/project/pandas/> та **excel-writer** <https://pypi.org/project/excel-writer/>.

Таблиці частот букв та біграм тексту з файлу FILE

Монограми без пробілів

Monogram	Frequency
о	0.11502
е	0.08765
а	0.08307
н	0.06936
и	0.06641
т	0.05995
с	0.05312
л	0.0504
в	0.04713
р	0.03975
к	0.0344
д	0.02958
м	0.02871
у	0.02713
п	0.0242
я	0.0216
ь	0.02011
ы	0.01852
г	0.01822
б	0.01754
ч	0.01698
з	0.01632
ж	0.01135
й	0.01041
ш	0.00862
х	0.00774
ю	0.00627
э	0.00356
щ	0.00285
ц	0.0028
ф	0.00125

Монограми з пробілами

Monogram	Frequency
	0.167557315
о	0.095748412
е	0.072960338
а	0.069150495
н	0.057736605
и	0.055286346
т	0.049904558
с	0.044221376
л	0.041950991
в	0.039230018
р	0.03309324
к	0.028632481
д	0.024619903
м	0.023901008
у	0.022584738
п	0.020142299
я	0.017982004
ь	0.016739729
ы	0.015420451
г	0.015170793
б	0.014599287
ч	0.014133058
з	0.013586818
ж	0.009446102
й	0.00866404
ш	0.007171506
х	0.006442383
ю	0.00521575
э	0.002963412
щ	0.002372655
ц	0.00233295
ф	0.001038939

Біграми, що не перетинаються:

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с
а	0.00048	0.00149	0.00554	0.00115	0.0036	0.00212	0.0015	0.00554	0.00162	0.00066	0.00644	0.01196	0.00351	0.00776	0.00205	0.00247	0.00322	0.0062
б	0.00099	5.8E-06	0.00012	1.2E-05	4E-05	0.00246	4.3E-06	5.8E-06	0.00093	0	0.0003	0.00094	7.7E-05	0.00037	0.00245	7.2E-06	0.00136	0.00012
в	0.00677	0.00019	0.00034	0.00034	0.00079	0.00547	5.8E-05	0.00074	0.00555	0	0.00071	0.00108	0.00036	0.00194	0.00886	0.00085	0.00148	0.00483
г	0.00104	4.6E-05	8.4E-05	1.3E-05	0.00141	0.00043	1.9E-05	2.7E-05	0.00091	0	0.00015	0.00181	1.6E-05	0.00026	0.00993	0.00011	0.00079	0.00012
д	0.00515	8.8E-05	0.00099	2.9E-05	8.5E-05	0.00534	2.3E-05	4.8E-05	0.00254	0	0.00033	0.00098	0.00013	0.00192	0.00433	0.00016	0.00174	0.00056
е	0.00037	0.00257	0.00501	0.00488	0.00394	0.00312	0.00126	0.00218	0.00156	0.00292	0.00308	0.00739	0.00611	0.01049	0.00191	0.00365	0.00674	0.00667
ж	0.00163	6.6E-05	2.5E-05	7.2E-06	0.00095	0.00488	1.3E-05	1.3E-05	0.00173	0	0.00012	1E-05	1.3E-05	0.00121	7.8E-05	3.8E-05	7.2E-06	0.00017
з	0.00631	0.0002	0.00105	0.00063	0.00091	0.00034	0.00015	0.00011	0.00039	0	0.00023	0.00021	0.00039	0.00236	0.00051	0.00014	0.00028	0.00019
и	0.00034	0.00137	0.0053	0.00094	0.00285	0.00344	0.00064	0.00265	0.00194	0.00158	0.00307	0.00576	0.00407	0.00691	0.00202	0.00235	0.00113	0.00465
й	0.00043	0.0003	0.00071	0.00023	0.00073	0.00012	0.00018	0.0002	0.00079	0	0.00061	0.0003	0.00035	0.00092	0.00056	0.00079	0.00031	0.00141
к	0.0092	0.00039	0.00058	9.5E-05	0.00024	0.00071	0.00024	0.0001	0.00367	0	0.00035	0.00048	0.0002	0.00144	0.00972	0.00032	0.00175	0.00158
л	0.00933	0.00022	0.00083	0.00028	0.00028	0.00668	0.00047	0.00014	0.00678	0	0.00085	0.00063	0.00016	0.00104	0.00808	0.0006	0.00015	0.00236
м	0.00306	0.00034	0.0007	0.00027	0.00037	0.00356	0.00014	0.00018	0.00373	0	0.00055	0.00032	0.00028	0.00217	0.00435	0.00078	0.00027	0.00094
н	0.01332	0.00035	0.00068	0.00026	0.00099	0.01189	6.5E-05	0.00022	0.00952	0	0.00049	0.0001	0.00014	0.00414	0.01206	0.00073	0.00021	0.00191
о	0.00029	0.00537	0.01131	0.0061	0.00611	0.00371	0.00286	0.00181	0.00225	0.00346	0.00307	0.00725	0.00665	0.01282	0.00299	0.0036	0.00703	0.00961
п	0.0014	1E-05	2.9E-06	0	2.9E-06	0.0025	0	1.4E-06	0.00096	0	8.2E-05	0.00054	1.4E-06	8.8E-05	0.00963	2.9E-05	0.00737	1.4E-05
р	0.00821	0.00031	0.00045	0.00046	0.00042	0.00616	0.00036	3.3E-05	0.00532	0	0.00075	7.1E-05	0.00017	0.00087	0.00813	0.00015	3E-05	0.00037
с	0.00214	0.00016	0.00239	0.00012	0.00057	0.00531	0.00022	8.8E-05	0.00169	0	0.00634	0.00311	0.00108	0.00162	0.00309	0.00226	0.00024	0.00097
т	0.00607	0.00028	0.00035	7.1E-05	0.00037	0.00615	7.8E-05	0.00011	0.00487	0	0.0008	0.00042	0.0002	0.00174	0.01735	0.00042	0.00302	0.00153
у	0.00021	0.00083	0.00194	0.00125	0.00222	0.00046	0.00207	0.00053	0.00077	0.00014	0.00141	0.0015	0.00166	0.00123	0.00051	0.00135	0.00077	0.00181
ф	0.00012	2.9E-06	1E-05	0	0	0.00014	0	0	0.0004	0	1.4E-06	2.5E-05	5.8E-06	5.8E-06	1E-04	1.4E-06	0.00029	1.7E-05
х	0.00101	0.00013	0.00037	9.3E-05	0.00019	0.00011	5.8E-05	6.2E-05	0.00038	0	0.00017	0.00022	0.00018	0.00038	0.0029	0.00033	0.00018	0.00039
ц	0.00057	1.4E-05	6.4E-05	8.7E-06	1.3E-05	0.00081	2.9E-06	8.7E-06	0.00023	0	0.00013	0	1.4E-05	3.8E-05	0.00036	3.5E-05	8.7E-06	2.6E-05
ч	0.00264	2.6E-05	0.00012	1.3E-05	3.8E-05	0.00397	2.9E-06	3.5E-05	0.00157	0	0.00039	5.3E-05	2.2E-05	0.00065	0.00012	9.3E-05	9.4E-05	0.00011
ш	0.00109	4.3E-06	9.8E-05	1.4E-06	0	0.00268	1.4E-06	2.9E-06	0.00185	0	0.00043	0.00062	1.2E-05	0.00042	0.00033	1E-05	4.3E-06	7.2E-06
щ	0.00035	0	0	0	0	0.00151	0	0	0.00078	0	0	0	0	3.2E-05	1.4E-06	1.4E-06	7.2E-06	0
ы	7.8E-05	0.00081	0.00162	0.00021	0.00037	0.00133	0.00011	0.00027	0.00043	0.00138	0.00043	0.00311	0.00152	0.00101	0.00031	0.00063	0.00043	0.00143
ь	0.00018	0.00042	0.00119	0.00024	0.00053	0.00148	9.5E-05	0.00055	0.00145	0	0.00215	0.00023	0.00066	0.00261	0.00097	0.00099	0.00028	0.00224
э	0	0	2.9E-06	2.9E-06	4.3E-06	0	0	8.7E-06	0	0	5.3E-05	5.1E-05	0	3E-05	0	1.2E-05	8.7E-06	5.8E-06
ю	0.0001	0.00088	0.00026	9.3E-05	0.00043	0.0001	0.00013	9.1E-05	0.00033	0	0.00025	0.00012	0.00016	0.00032	0.00032	0.00028	0.00019	0.00056
я	0.00037	0.00052	0.00152	0.00063	0.0012	0.00056	0.00049	0.00068	0.00112	0.00013	0.0009	0.00122	0.00076	0.00262	0.00082	0.00108	0.00033	0.00233

т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я
0.00593	0.00059	0.00036	0.00117	0.00017	0.00165	0.0009	0.00024	0	0	0.0004	0.00103	0.00312
1.3E-05	0.00118	0	6.9E-05	4.3E-06	1E-05	1.3E-05	0.00031	0.00498	0.00015	0.00011	4.3E-06	0.00065
0.0007	0.00087	1.4E-05	0.0001	4.3E-05	0.0003	0.00132	1.2E-05	0.00284	0.00021	0.00029	2.9E-06	0.00035
2.6E-05	0.00072	2.9E-06	4.3E-06	0	6.2E-05	8.7E-06	0	0	0	7.2E-06	5.8E-06	8.7E-06
0.00029	0.00219	1.4E-06	8.2E-05	0.00021	8.8E-05	0.00012	0	0.00059	0.00097	2.2E-05	5.8E-06	0.00048
0.00657	0.00072	7.1E-05	0.00122	0.00022	0.00167	0.00122	0.00074	0	0	0.00034	0.00049	0.00064
3.3E-05	0.0003	0	1.4E-06	1.4E-06	6.8E-05	1.4E-06	0	0	4.8E-05	8.7E-06	2.9E-06	2E-05
0.0001	0.00036	5.8E-06	4.3E-06	8.7E-06	9.5E-05	1.9E-05	0	0.0004	0.00021	3.6E-05	8.7E-06	0.00056
0.00561	0.00064	8.5E-05	0.00171	0.001	0.00292	0.00057	0.0002	0	0	0.0003	0.00031	0.00239
0.0006	0.00018	4.8E-05	8.5E-05	8.1E-05	0.00034	0.00013	8.7E-06	0	0	6.9E-05	2.9E-06	6.5E-05
0.00072	0.00176	2.3E-05	8.1E-05	1.3E-05	0.00016	3.2E-05	7.2E-06	0	0	0.00016	0	0.00012
0.00041	0.00158	2E-05	4.6E-05	2E-05	0.00067	4.2E-05	2.9E-06	0.00111	0.00407	0.00016	0.00119	0.00207
0.00039	0.00349	2.7E-05	9.3E-05	2.2E-05	0.00055	7.2E-05	1.2E-05	0.00111	7.9E-05	7.5E-05	2.9E-06	0.00058
0.00075	0.00315	3E-05	0.0001	0.00043	0.00048	2.3E-05	0.00019	0.00332	0.00144	4.5E-05	0.00018	0.00211
0.00943	0.00077	0.00026	0.00065	0.00011	0.00332	0.00151	0.00018	0	0	0.00056	0.00095	0.00115
5.6E-05	0.00062	0	0	1.7E-05	4.6E-05	1.2E-05	1.4E-06	0.0003	3.3E-05	0	4.3E-06	0.00043
0.00057	0.00275	2.3E-05	0.00011	3.8E-05	9.7E-05	0.00025	4.6E-05	0.00146	0.00101	2.9E-06	0.00014	0.001
0.01268	0.00078	1.4E-05	0.0002	6.5E-05	0.00056	8.7E-05	0	0.00039	0.00328	7.1E-05	0.00015	0.00344
0.00036	0.0015	5.3E-05	5.8E-05	8.8E-05	0.00053	1.3E-05	2.2E-05	0.00177	0.00744	0.00014	5.8E-05	0.00071
0.00158	0.00013	6.8E-05	0.00044	7.2E-06	0.00158	0.00081	0.00023	0	0	9E-05	0.0012	0.00024
1.9E-05	3E-05	8.7E-06	0	0	2.9E-06	0	0	4.3E-06	6.1E-05	0	0	1.4E-06
0.00012	0.00026	1.3E-05	2.2E-05	1.2E-05	7.7E-05	2.2E-05	1.4E-06	0	0	3.8E-05	0	3E-05
1.4E-05	0.00024	0	0	7.2E-06	8.7E-06	2.9E-06	0	0.00018	0	1.2E-05	0	4.3E-06
0.00527	0.00112	0	1.4E-05	0	1.7E-05	0.00021	0	0	0.00019	7.2E-06	0	1.2E-05
8.8E-05	0.00035	2.9E-06	4.3E-06	2.9E-06	4.3E-06	2.9E-06	0	0	0.00055	0	5.8E-06	0
0	0.0001	0	0	0	0	0	0	0	2.2E-05	0	0	0
0.00105	0.00012	1.4E-05	0.00098	4.3E-06	0.00034	0.0005	4.3E-06	0	0	5.6E-05	1.4E-06	5.6E-05
0.00068	0.00031	3.2E-05	0.0001	0.00015	0.0007	0.00056	2.6E-05	0	0	0.00033	0.00048	0.00085
0.00336	0	1.4E-06	0	0	0	0	0	0	0	0	0	0
0.00046	0.00013	3.3E-05	3E-05	2E-05	0.00023	7.7E-05	0.00041	0	0	6.1E-05	4.5E-05	5.9E-05
0.00226	0.00028	4.2E-05	0.00034	5.5E-05	0.00058	9.8E-05	0.00024	0	0	0.00019	0.0001	0.0003

Біграми, що не перетинаються, враховуючи пробіли:

		а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с
	0	0.00318	0.00673	0.01475	0.00342	0.00724	0.00537	0.00249	0.00414	0.01102	2.4E-06	0.00908	0.00361	0.00554	0.01654	0.01355	0.01402	0.00379	0.01602
а	0.01875	4.8E-06	0.00062	0.00304	0.00059	0.00215	0.00099	0.00111	0.00407	0.00013	0.00054	0.00419	0.00957	0.0024	0.00462	7.2E-06	0.00058	0.00227	0.0035
б	0.00041	0.00083	2.4E-06	7.6E-05	3.6E-06	3.5E-05	0.00195	3.6E-06	1.2E-06	0.00072	0	0.00025	0.00074	5.8E-05	0.00028	0.00193	0	0.0011	6.7E-05
в	0.00531	0.00538	1.4E-05	1.9E-05	3.4E-05	0.00032	0.00419	2.4E-06	0.00055	0.00446	0	0.00015	0.00076	0.00015	0.00124	0.00702	0.00019	0.00101	0.00339
г	0.00068	0.00081	0	7.2E-06	3.6E-06	0.00116	0.00034	0	1.2E-06	0.00075	0	0.00013	0.00153	0	0.00017	0.00083	0	0.00067	3E-05
д	0.00096	0.00419	2.3E-05	0.00078	4.8E-06	4.6E-05	0.0044	1.6E-05	0	0.00204	0	0.00019	0.0008	7E-05	0.00152	0.00365	6.5E-05	0.00148	0.0003
е	0.01893	4.8E-05	0.00125	0.00246	0.00361	0.00242	0.00215	0.00076	0.00112	9.1E-05	0.00245	0.00176	0.00565	0.0039	0.00708	0.00018	0.00138	0.0052	0.00377
ж	0.00027	0.00139	5.5E-05	0	6E-06	0.00081	0.00399	8.4E-06	0	0.00148	0	9.5E-05	2.4E-06	2.4E-06	0.00098	4.1E-05	0	0	9.7E-05
з	0.00108	0.00536	0.00017	0.00082	0.0005	0.00077	0.00024	0.00012	1.7E-05	0.00031	0	0.0001	0.00018	0.00027	0.0019	0.00039	0	0.00021	6E-05
и	0.01827	8.3E-05	0.00039	0.0026	0.00041	0.00151	0.00242	0.00031	0.00167	0.00057	0.00128	0.00153	0.00442	0.00279	0.00403	8.1E-05	0.00013	0.00052	0.00212
й	0.00713	0	0	0	0	0.0002	0	0	1.2E-06	0	0	6.6E-05	7.6E-05	4E-05	0.00026	1.2E-06	0	2.4E-06	0.00047
к	0.00464	0.00769	0	0.00017	0	4.8E-06	0.00048	1.2E-05	2.4E-06	0.00289	0	1.8E-05	0.00032	0	0.00059	0.00768	0	0.00138	0.00091
л	0.00722	0.00759	1.8E-05	2.4E-06	9.7E-05	1.4E-05	0.00515	0.00038	2.4E-06	0.00522	0	0.00029	0.00027	1.2E-06	0.00028	0.00573	3.7E-05	0	0.00117
м	0.007	0.0024	6E-06	1.2E-06	2.6E-05	0	0.00295	1.2E-06	1.2E-06	0.00251	0	5.8E-05	8.8E-05	4.2E-05	0.00127	0.00312	5.8E-05	7.9E-05	0.00011
н	0.00543	0.01076	1.3E-05	9.6E-06	0.0001	0.00063	0.0098	4.8E-06	1.6E-05	0.00755	0	0.00024	0	0	0.0028	0.00963	0	7.3E-05	0.00109
о	0.02395	1.2E-06	0.00334	0.00734	0.00461	0.00414	0.0025	0.0018	0.00096	0.00071	0.003	0.00148	0.00546	0.00475	0.00823	0.00017	0.00112	0.00531	0.00574
п	5.2E-05	0.00115	4.8E-06	0	0	0	0.00208	0	0	0.00079	0	6.1E-05	0.00045	0	6.9E-05	0.00811	2.2E-05	0.00615	4.8E-06
р	0.00064	0.00677	0.00023	0.0003	0.00035	0.00035	0.00515	0.00028	1.2E-05	0.00428	0	0.00057	5.2E-05	0.00012	0.00068	0.00682	5.8E-05	1.7E-05	0.0002
с	0.00291	0.00177	5.2E-05	0.00179	1.3E-05	0.0003	0.00443	2.3E-05	1.3E-05	0.00131	0	0.00489	0.00256	0.00079	0.00094	0.00239	0.00172	0.00013	0.00059
т	0.00434	0.00481	1.6E-05	0.00257	7.2E-06	0.00011	0.00494	3.6E-06	1.2E-06	0.00376	0	0.00045	0.00026	1.7E-05	0.00105	0.01417	4.9E-05	0.00247	0.00078
у	0.00664	3E-05	0.00051	0.00109	0.00098	0.0015	0.00027	0.0017	0.00028	2.4E-06	0.00011	0.00076	0.00111	0.00115	0.0002	3.6E-06	0.00054	0.00055	0.00101
ф	4E-05	7.6E-05	0	0	0	0	0.0001	0	0	0.0003	0	0	3.2E-05	2.4E-06	1.2E-06	9.5E-05	0	0.00027	8.4E-06
х	0.00269	0.00082	0	9.1E-05	2.4E-06	1.2E-06	3.5E-05	0	0	0.00011	0	0	7.3E-05	5.1E-05	8.8E-05	0.00216	0	6.1E-05	4.7E-05
ц	0.00027	0.00046	1.2E-06	4E-05	0	0	0.00066	0	0	0.00018	0	0.00011	0	0	0	0.00028	0	0	0
ч	0.00078	0.00219	0	6E-06	0	0	0.00329	0	0	0.00132	0	0.00031	3.9E-05	9.6E-06	0.00045	5.4E-05	0	5.4E-05	0
ш	6.3E-05	0.00089	0	7.6E-05	0	0	0.00227	0	0	0.00148	0	0.00036	0.00055	6E-06	0.00037	0.00025	7.2E-06	2.4E-06	0
щ	2.4E-06	0.0003	0	0	0	0	0.00129	0	0	0.00066	0	0	0	0	2.9E-05	1.2E-06	0	2.4E-06	0
ы	0.00417	0	0.00047	0.00094	8.4E-05	0.00014	0.00094	2.9E-05	5.9E-05	4.8E-06	0.00118	0.00016	0.00252	0.00116	0.00033	0	0.00011	0.00029	0.00074
ь	0.01043	0	7.9E-05	2.8E-05	4.1E-05	2.4E-05	0.0006	2.4E-06	0.00017	0.00042	0	0.00117	0	0.00023	0.00112	9.6E-06	4.8E-06	0	0.00083
э	0	0	0	0	1.2E-06	1.2E-06	0	0	4.8E-06	0	2.4E-06	3E-05	4.9E-05	1.2E-06	2.3E-05	0	7.2E-06	8.4E-06	7.2E-06
ю	0.00345	2.4E-06	0.00063	2.4E-06	2.4E-06	0.00024	0	7.2E-06	3.6E-06	1.2E-06	0	3.6E-06	1.3E-05	1.3E-05	2.4E-05	0	9.6E-06	3.6E-05	0.00015
я	0.01106	0	2.2E-05	0.00026	0.00026	0.00053	8.3E-05	0.00028	0.00029	4.2E-05	0.00012	0.00012	0.00081	0.00032	0.00066	0	4.8E-05	5.1E-05	0.00088

т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я
0.00838	0.00411	0.00051	0.0016	0.00025	0.00621	0.00072	9.6E-05	0	0	0.00293	3.5E-05	0.00221
0.00411	5.9E-05	0.00027	0.00084	0.00011	0.00077	0.00064	0.0002	0	0	6E-06	0.00084	0.00236
1.1E-05	0.00098	0	6.1E-05	7.2E-06	9.6E-06	8.4E-06	0.00023	0.00411	0.00012	0	6E-06	0.00054
0.00022	0.00066	0	4.9E-05	1.9E-05	8.2E-05	0.0011	1.1E-05	0.00242	0.00019	0	2.4E-06	0.00026
6E-06	0.00056	0	0	0	4.1E-05	4.8E-06	0	0	0	0	3.6E-06	0
0.00018	0.00181	0	4.9E-05	0.00016	3.1E-05	0.00011	0	0.0005	0.00082	0	9.6E-06	0.0004
0.00449	9.7E-05	6E-06	0.00075	0.00016	0.00092	0.0009	0.00059	0	0	0	0.0004	0.00034
0	0.00023	0	0	0	3.9E-05	0	0	0	4.7E-05	0	1.2E-06	0
1.4E-05	0.00023	0	0	6E-06	5.4E-05	2.4E-06	0	0.00037	0.00017	0	4.8E-06	0.00045
0.00379	8.4E-06	1.6E-05	0.0013	0.00079	0.00187	0.00041	0.00016	0	0	1.2E-06	0.00028	0.00174
0.00023	0	4.8E-06	0	4.8E-05	7.1E-05	5.1E-05	1.2E-06	0	0	0	0	0
0.00038	0.00143	2.4E-06	2.4E-06	6E-06	0	1.2E-05	0	0	0	0	0	0
1.7E-05	0.00117	6E-06	0	1.2E-06	0.00013	3.6E-06	1.2E-06	0.0009	0.00355	0	0.00098	0.0017
1.3E-05	0.00279	1.2E-06	0	1.2E-05	2.3E-05	2.4E-06	4.8E-06	0.0009	6.9E-05	0	3.6E-06	0.00037
0.00041	0.00242	1.3E-05	1.2E-06	0.00033	0.00024	2.4E-06	0.00016	0.00283	0.00127	0	0.00016	0.00168
0.00631	4.2E-05	0.00014	0.00031	8.9E-05	0.00157	0.00122	0.00013	0	0	3.1E-05	0.00076	0.00046
4.2E-05	0.00047	1.2E-06	0	1.1E-05	3.4E-05	8.4E-06	1.2E-06	0.00026	4E-05	0	2.4E-06	0.00032
0.00044	0.00222	1.3E-05	9.4E-05	2.5E-05	6.9E-05	0.00021	3.1E-05	0.0012	0.00085	0	0.00012	0.00088
0.01035	0.00052	8.4E-06	0.00012	6E-05	0.00041	6.5E-05	0	0.00031	0.0028	1.2E-06	0.00014	0.00295
8.5E-05	0.00112	1.9E-05	1.3E-05	4.2E-05	0.00026	0	1.1E-05	0.00144	0.0062	1.2E-06	4.8E-05	0.00052
0.00105	1.2E-06	1.9E-05	0.0003	2.4E-06	0.00086	0.00068	0.00021	0	0	1.1E-05	0.00092	0.00011
1.8E-05	2.9E-05	8.4E-06	0	0	0	0	0	3.6E-06	5.1E-05	0	0	2.4E-06
4.8E-06	0.00013	0	0	0	0	1.2E-05	0	0	1.2E-06	0	0	0
0	0.00022	0	0	6E-06	0	0	0	0.00017	0	0	0	0
0.00443	0.00091	0	0	0	0	0.00017	0	0	0.00016	0	0	0
6.6E-05	0.00027	0	0	2.4E-06	0	0	0	0	0.00048	0	3.6E-06	0
0	8.1E-05	0	0	0	0	0	0	0	2E-05	0	0	0
0.0007	8.4E-06	0	0.00081	0	0.00018	0.00039	1.2E-06	0	0	0	0	2.4E-06
6.5E-05	0	2.4E-06	1.2E-06	0.0001	5.7E-05	0.0004	1.3E-05	0	0	2.4E-06	0.00036	0.00049
0.00281	0	0	0	0	0	0	0	0	0	0	0	0
0.00024	0	1.2E-06	8.4E-06	6E-06	3.2E-05	2.5E-05	0.00031	0	0	0	3.1E-05	0
0.00138	0	0	0.00019	3.4E-05	0.00013	5.1E-05	0.00019	0	0	0	8.4E-05	0.00015

Біграми, що перетинаються:

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с
а	0.00053	0.00153	0.00569	0.00116	0.00357	0.00212	0.00152	0.00544	0.00163	0.00063	0.00637	0.01201	0.00352	0.00772	0.00204	0.00248	0.0032	0.00627
б	0.00102	3.6E-06	0.00011	9.4E-06	4.1E-05	0.00241	4.3E-06	5.1E-06	0.00097	0	0.00029	0.00094	7.2E-05	0.00037	0.00248	6.5E-06	0.0013	0.0001
в	0.00658	0.00018	0.00033	0.00033	0.00077	0.00532	6.3E-05	0.00074	0.00563	0	0.00071	0.00105	0.00039	0.00195	0.00879	0.00086	0.00146	0.00487
г	0.00101	4.1E-05	0.0001	1.4E-05	0.00143	0.00045	1.4E-05	2.7E-05	0.00094	0	0.00017	0.0018	1.9E-05	0.00029	0.01001	0.00012	0.00083	0.00013
д	0.00506	7.5E-05	0.00099	3E-05	8.2E-05	0.00539	3E-05	4E-05	0.00251	0	0.00032	0.00096	0.00011	0.00195	0.00446	0.00015	0.0018	0.00054
е	0.00034	0.00262	0.005	0.00482	0.00399	0.00309	0.00123	0.00218	0.00161	0.00297	0.00304	0.00739	0.00607	0.01041	0.00188	0.00369	0.00678	0.00672
ж	0.00165	6.6E-05	2.2E-05	7.9E-06	0.00093	0.00484	1.5E-05	1E-05	0.00173	0	0.00012	8.7E-06	1.2E-05	0.00122	7.8E-05	3.4E-05	4.3E-06	0.00014
з	0.00639	0.00021	0.00104	0.00066	0.00096	0.00032	0.00014	9.4E-05	0.0004	0	0.00021	0.00024	0.0004	0.00236	0.00047	0.00014	0.00027	0.00019
и	0.00034	0.00135	0.00537	0.00094	0.00279	0.00347	0.0006	0.00266	0.00194	0.00158	0.00303	0.00572	0.00395	0.00693	0.00198	0.00233	0.00113	0.00464
й	0.00042	0.0003	0.00071	0.00022	0.00073	0.00013	0.00017	0.00018	0.00079	0	0.00059	0.00029	0.00035	0.00094	0.00056	0.00074	0.00031	0.00143
к	0.00928	0.00039	0.00058	9.5E-05	0.00024	0.00072	0.00025	0.0001	0.00372	0	0.00033	0.0005	0.0002	0.00145	0.00974	0.00035	0.00174	0.00161
л	0.00926	0.00022	0.00085	0.00029	0.00026	0.0068	0.00049	0.00014	0.00677	0	0.00083	0.00064	0.00017	0.00099	0.00805	0.0006	0.00015	0.00232
м	0.00306	0.00031	0.0007	0.00026	0.00038	0.00362	0.00014	0.00019	0.0038	0	0.00056	0.00029	0.00028	0.00222	0.00444	0.00079	0.00027	0.00095
н	0.01339	0.00036	0.00065	0.00024	0.00101	0.01189	6.9E-05	0.00024	0.00949	0	0.00052	9.8E-05	0.00014	0.00415	0.01189	0.00074	0.00022	0.00194
о	0.0003	0.0055	0.01134	0.00596	0.00612	0.00384	0.00286	0.00184	0.00233	0.00354	0.00303	0.00706	0.00657	0.01279	0.00294	0.00357	0.00705	0.00953
п	0.00137	1.1E-05	4.3E-06	0	1.4E-06	0.00249	1.4E-06	1.4E-06	0.00096	0	7.4E-05	0.00055	1.4E-06	7.9E-05	0.00974	3E-05	0.00738	1.5E-05
р	0.00817	0.0003	0.00041	0.00045	0.00043	0.00619	0.00036	3.8E-05	0.00534	0	0.00074	6.9E-05	0.00016	0.00086	0.00821	0.00014	3.2E-05	0.00034
с	0.00218	0.00017	0.00239	0.00011	0.00056	0.00531	0.00023	8.7E-05	0.00168	0	0.00629	0.00308	0.0011	0.00159	0.00303	0.00225	0.00023	0.00098
т	0.00601	0.0003	0.00361	8E-05	0.00037	0.00618	1E-04	0.00011	0.00489	0	0.00083	0.00042	0.00021	0.00179	0.01739	0.00044	0.00301	0.00149
у	0.00022	0.00084	0.00197	0.00128	0.00216	0.00046	0.00214	0.00052	0.00075	0.00014	0.00142	0.0015	0.00167	0.00123	0.00053	0.00128	0.0008	0.00188
ф	0.00011	3.6E-06	1E-05	7.2E-07	0	0.00014	0	0	0.00037	0	2.2E-06	3E-05	6.5E-06	5.1E-06	0.0001	2.2E-06	0.0003	1.6E-05
х	0.00102	0.00013	0.00037	9E-05	0.00019	0.00011	5.3E-05	7E-05	0.0004	0	0.00018	0.00023	0.00018	0.00037	0.00291	0.00031	0.00017	0.00038
ц	0.00056	1.2E-05	7.9E-05	8.7E-06	1.2E-05	0.00079	2.9E-06	7.9E-06	0.00022	0	0.00015	1.4E-06	1.1E-05	2.8E-05	0.00036	3.3E-05	6.5E-06	3E-05
ч	0.00265	2.7E-05	0.00013	1.4E-05	3.7E-05	0.00401	5.8E-06	3.8E-05	0.00159	0	0.00041	6.4E-05	1.8E-05	0.00067	0.00012	0.00012	9.2E-05	0.0001
ш	0.00108	6.5E-06	0.00011	2.2E-06	7.2E-07	0.00269	7.2E-07	1.4E-06	0.00182	0	0.00043	0.00064	1.2E-05	0.00044	0.00034	1.7E-05	4.3E-06	5.1E-06
щ	0.00034	0	7.2E-07	0	0	0.00155	0	0	0.00078	0	0	1.4E-06	7.2E-07	3.5E-05	2.2E-06	2.2E-06	6.5E-06	1.4E-06
ы	7.7E-05	0.00081	0.0016	0.00021	0.0004	0.00125	9.3E-05	0.00025	0.00043	0.0014	0.00044	0.00314	0.00155	0.00102	0.0003	0.00061	0.00041	0.00139
ь	0.00017	0.00042	0.0012	0.00023	0.00051	0.00142	0.00011	0.00051	0.00144	0	0.00215	0.00023	0.00063	0.00254	0.00098	0.00097	0.00027	0.00226
э	0	7.2E-07	2.2E-06	2.9E-06	2.2E-06	0	0	5.8E-06	0	3.6E-06	4.7E-05	5.1E-05	2.2E-06	3E-05	0	8.7E-06	8.7E-06	4.3E-06
ю	0.0001	0.00084	0.00027	9.6E-05	0.00045	0.00011	0.00013	9.6E-05	0.00035	0	0.00026	0.00012	0.00015	0.00033	0.00033	0.00031	0.0002	0.00057
я	0.00034	0.0005	0.00148	0.0006	0.00117	0.00054	0.00053	0.0007	0.00111	0.00014	0.00087	0.00124	0.00076	0.00261	0.00085	0.00109	0.00033	0.00229

т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я
0.006	0.0006	0.00036	0.00116	0.00017	0.00164	0.00085	0.00025	0	0	0.00041	0.00101	0.00321
1.3E-05	0.00121	7.2E-07	6.5E-05	7.9E-06	8.7E-06	1.1E-05	0.00028	0.00492	0.00015	0.0001	5.8E-06	0.00062
0.00071	0.00089	1.7E-05	9.8E-05	4.6E-05	0.00028	0.00133	1E-05	0.00286	0.00023	0.00029	5.1E-06	0.00034
2.8E-05	0.00073	3.6E-06	3.6E-06	0	5.3E-05	8.7E-06	0	0	0	1E-05	4.3E-06	5.8E-06
0.00028	0.00217	4.3E-06	7.6E-05	0.00021	7.6E-05	0.00012	7.2E-07	0.00061	0.00099	2.5E-05	8.7E-06	0.00048
0.00642	0.00071	7.2E-05	0.00123	0.00022	0.00171	0.00119	0.00074	0	0	0.00037	0.00048	0.00066
3.1E-05	0.00029	0	7.2E-07	7.2E-07	5.2E-05	7.2E-07	0	0	4.9E-05	1E-05	1.4E-06	1.7E-05
9E-05	0.00035	4.3E-06	4.3E-06	1E-05	9E-05	1.7E-05	0	0.00044	0.00021	3.9E-05	5.1E-06	0.00057
0.00561	0.00063	8.9E-05	0.00174	0.00101	0.0029	0.00057	0.0002	0	0	0.00029	0.0003	0.00233
0.00059	0.00017	4.5E-05	8.2E-05	7.3E-05	0.00034	0.00011	1E-05	0	0	6.9E-05	2.2E-06	7.6E-05
0.00072	0.0018	1.6E-05	8.8E-05	1.5E-05	0.00018	3.2E-05	8.7E-06	0	0	0.00013	1.4E-06	0.00011
0.0004	0.00159	2.4E-05	4.3E-05	1.5E-05	0.00066	4.3E-05	4.3E-06	0.00111	0.00418	0.00016	0.00119	0.00211
0.00038	0.00348	2.7E-05	9.5E-05	2.1E-05	0.00053	7.1E-05	9.4E-06	0.00108	7.9E-05	8.2E-05	7.2E-06	0.00058
0.00074	0.00314	2.8E-05	9.6E-05	0.00042	0.00049	2.3E-05	0.00019	0.0034	0.00146	4.8E-05	0.00019	0.00209
0.00936	0.00076	0.00027	0.00064	0.00013	0.00332	0.00157	0.00018	0	0	0.00055	0.00095	0.00112
5.9E-05	0.0006	1.4E-06	0	1.7E-05	3.8E-05	1.6E-05	7.2E-07	0.0003	4.3E-05	1.4E-06	3.6E-06	0.0004
0.00057	0.00272	1.9E-05	0.00012	3.5E-05	9.3E-05	0.00025	3.8E-05	0.00145	0.00103	4.3E-06	0.00013	0.00106
0.01267	0.00075	1.8E-05	0.00018	6.9E-05	0.00055	8.2E-05	4.3E-06	0.00039	0.00337	6.1E-05	0.00015	0.00357
0.00035	0.00149	4.9E-05	5.3E-05	7.8E-05	0.00053	1.5E-05	2E-05	0.00177	0.00745	0.00015	5.8E-05	0.00071
0.00162	0.00013	6.1E-05	0.00041	1.2E-05	0.00155	0.00085	0.00023	0	0	8.8E-05	0.00113	0.00025
2.5E-05	3.1E-05	8.7E-06	0	0	2.2E-06	0	0	5.8E-06	6.4E-05	0	1.4E-06	1.4E-06
0.00011	0.00027	1.1E-05	2.2E-05	1.1E-05	6.2E-05	2.5E-05	2.2E-06	0	7.2E-07	3.2E-05	2.2E-06	3.3E-05
1.3E-05	0.00026	0	7.2E-07	1E-05	6.5E-06	1.4E-06	0	0.00018	0	8.7E-06	0	5.8E-06
0.00531	0.00112	0	9.4E-06	7.2E-07	1.7E-05	0.0002	0	0	0.00019	4.3E-06	0	2.2E-05
8.2E-05	0.00034	1.4E-06	3.6E-06	2.9E-06	2.9E-06	2.9E-06	0	0	0.00058	7.2E-07	4.3E-06	7.2E-07
0	0.0001	0	0	0	0	0	0	0	2.6E-05	7.2E-07	0	0
0.00105	0.0001	1.3E-05	0.00103	7.2E-06	0.00031	0.00051	3.6E-06	0	0	5.3E-05	7.2E-07	5.7E-05
0.00066	0.00029	3.5E-05	0.00011	0.00014	0.00068	0.00053	2.2E-05	0	0	0.00031	0.00047	0.00082
0.00339	0	2.2E-06	0	0	0	7.2E-07	0	0	0	0	0	0
0.00047	0.00013	3.3E-05	3.9E-05	2.2E-05	0.00024	6.8E-05	0.0004	0	0	6.3E-05	4.3E-05	5.6E-05
0.00221	0.00027	4E-05	0.00033	6.5E-05	0.00057	1E-04	0.00024	0	0	0.00019	0.00012	0.00029

Біграми, що перетинаються, враховуючи пробіли:

		а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с
	0	0.00317	0.00669	0.01476	0.00345	0.00733	0.00525	0.00252	0.00415	0.01105	2.4E-06	0.00902	0.00365	0.00556	0.01642	0.01351	0.01405	0.00382	0.01615
а	0.0187	6.6E-06	0.00059	0.00303	0.00058	0.00215	0.001	0.0011	0.00414	0.00013	0.00052	0.00413	0.00959	0.00239	0.00452	5.4E-06	0.00059	0.00229	0.00347
б	0.0004	0.00083	1.2E-06	8.6E-05	3.6E-06	3.2E-05	0.00199	3.6E-06	2.4E-06	0.00077	0	0.00024	0.00077	5.7E-05	0.0003	0.00195	0	0.00108	7.5E-05
в	0.00537	0.00541	1.6E-05	2.1E-05	3E-05	0.00031	0.00421	1.2E-06	0.00053	0.00449	0	0.00015	0.00077	0.00014	0.00125	0.00694	0.0002	0.00104	0.00343
г	0.00069	0.00083	0	7.2E-06	2.4E-06	0.00114	0.00035	0	6E-07	0.00074	0	0.00012	0.00149	6E-07	0.00017	0.00829	0	0.00067	2.9E-05
д	0.00095	0.00419	2.2E-05	0.00075	4.2E-06	3.7E-05	0.00446	1.4E-05	6E-07	0.00202	0	0.00021	0.00078	7E-05	0.0015	0.00366	6.4E-05	0.00147	0.00033
е	0.01888	4.8E-05	0.00128	0.00247	0.00363	0.00241	0.00216	0.00079	0.00114	9.6E-05	0.00247	0.00174	0.00569	0.00395	0.00714	0.00018	0.00134	0.00516	0.00378
ж	0.00028	0.00137	4.3E-05	0	5.4E-06	0.00075	0.00402	1.1E-05	0	0.00142	0	9E-05	3.6E-06	2.4E-06	0.00098	4.5E-05	0	6E-07	9.6E-05
з	0.00105	0.00531	0.00016	0.00079	0.00052	0.00074	0.00024	0.0001	1.9E-05	0.0003	0	8.8E-05	0.00018	0.00027	0.00189	0.00035	0	0.00019	5.2E-05
и	0.01812	8.5E-05	0.00038	0.00263	0.00041	0.00153	0.00237	0.00031	0.00172	0.00059	0.00132	0.00156	0.00439	0.00275	0.00405	8.8E-05	0.00014	0.00051	0.0021
й	0.00711	0	1.2E-06	0	0	0.0002	1.2E-06	0	6E-07	0	0	6.4E-05	8.6E-05	5.2E-05	0.00025	1.8E-06	0	1.2E-06	0.00049
к	0.00468	0.00765	0	0.00016	0	3.6E-06	0.0005	1.1E-05	2.4E-06	0.00284	0	1.7E-05	0.00032	6E-07	0.00061	0.00773	0	0.00136	0.00093
л	0.00722	0.00759	1.7E-05	3E-06	0.00012	1.4E-05	0.0052	0.00037	3E-06	0.00521	0	0.00028	0.00025	6E-07	0.00029	0.00575	4E-05	0	0.00115
м	0.007	0.00244	1.3E-05	1.8E-06	2.9E-05	0	0.00291	6E-07	3E-06	0.00261	0	5.2E-05	9E-05	4E-05	0.00124	0.00312	5.5E-05	7.2E-05	0.0001
н	0.00547	0.01084	1.3E-05	7.8E-06	9.9E-05	0.00064	0.00981	4.2E-06	1.4E-05	0.00759	0	0.00025	0	1.8E-06	0.0028	0.00959	0	7.9E-05	0.00103
о	0.02401	2.4E-06	0.00337	0.00725	0.00454	0.00414	0.00244	0.00188	0.00099	0.00071	0.00295	0.00152	0.00542	0.00475	0.00825	0.00018	0.0011	0.00536	0.00578
п	4.5E-05	0.00114	7.2E-06	0	0	0	0.00207	0	0	0.00079	0	5.8E-05	0.00045	0	6.3E-05	0.0081	2.4E-05	0.00615	6E-06
р	0.00066	0.00679	0.00022	0.00029	0.00036	0.00034	0.00514	0.00029	1.6E-05	0.00437	0	0.00056	4.9E-05	0.00012	0.00066	0.00678	5.8E-05	1.6E-05	0.0002
с	0.00289	0.00175	5.2E-05	0.00179	1E-05	0.00032	0.00436	2.6E-05	1.9E-05	0.00131	0	0.00501	0.00249	0.00081	0.00094	0.00237	0.00167	0.00012	0.00059
т	0.00437	0.00492	1.2E-05	0.00261	6E-06	0.00012	0.00496	4.8E-06	1.8E-06	0.00378	0	0.00047	0.00025	1.7E-05	0.00105	0.01421	4.8E-05	0.00241	0.00083
у	0.00664	3.5E-05	0.00051	0.0011	0.00096	0.0015	0.00025	0.00169	0.00029	3E-06	0.00012	0.00075	0.00112	0.00116	0.00021	3.6E-06	0.00056	0.00054	0.00098
ф	4E-05	8.6E-05	0	0	0	0	0.00012	0	0	0.00031	0	0	2.5E-05	3E-06	1.2E-06	8.4E-05	0	0.00025	9.6E-06
х	0.00271	0.00081	0	9.4E-05	1.2E-06	6E-07	3.4E-05	0	0	0.00011	0	0	8.5E-05	4.9E-05	9.7E-05	0.0022	6E-07	6.3E-05	4.1E-05
ц	0.00026	0.00046	6E-07	3.8E-05	0	0	0.00065	0	0	0.00016	0	0.00011	6E-07	1.2E-06	0	0.00027	0	0	0
ч	0.00078	0.0022	0	3.6E-06	0	0	0.00333	0	0	0.00128	0	0.00031	4.3E-05	4.8E-06	0.00048	4.8E-05	0	6E-05	0
ш	6.4E-05	0.0009	0	8.4E-05	0	0	0.00224	0	0	0.00151	0	0.00035	0.00053	4.8E-06	0.00036	0.00028	6E-06	1.2E-06	0
щ	6.6E-06	0.00028	0	0	0	0	0.00129	0	0	0.00065	0	0	0	0	2.9E-05	1.2E-06	0	5.4E-06	0
ы	0.00421	0	0.00046	0.00095	8.8E-05	0.00013	0.00094	2.5E-05	6E-05	5.4E-06	0.00117	0.00016	0.00254	0.00114	0.00032	0	0.00011	0.00028	0.00074
ь	0.01044	0	8.7E-05	2.8E-05	4.9E-05	2.6E-05	0.0006	1.2E-06	0.00017	0.00041	0	0.00117	0	0.00021	0.00115	8.4E-06	3.6E-06	0	0.00082
э	2.4E-06	0	0	1.2E-06	2.4E-06	1.8E-06	0	0	4.8E-06	0	3E-06	3.9E-05	4.3E-05	1.8E-06	2.3E-05	0	7.2E-06	7.2E-06	3.6E-06
ю	0.00339	1.8E-06	0.00062	3E-06	3.6E-06	0.00023	0	3.6E-06	3E-06	1.2E-06	0	3.6E-06	1.4E-05	1.6E-05	2.5E-05	0	8.4E-06	3.9E-05	0.00016
я	0.01112	0	2.5E-05	0.00025	0.00028	0.00053	7E-05	0.00028	0.00031	3.7E-05	0.00012	0.00011	0.0008	0.00031	0.00065	0	5.5E-05	4.5E-05	0.00084

т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я
0.0083	0.00415	0.00053	0.00153	0.00025	0.00629	0.0007	0.0001	6E-07	6E-07	0.0029	3.4E-05	0.00222
0.00414	5.1E-05	0.00025	0.00083	0.00012	0.00077	0.00063	0.0002	0	0	6E-06	0.00084	0.00238
7.8E-06	0.00097	0	5.4E-05	6E-06	6E-06	8.4E-06	0.00023	0.00409	0.00013	0	4.8E-06	0.0005
0.00022	0.00065	0	5.5E-05	2.3E-05	7.8E-05	0.00106	7.8E-06	0.00238	0.00019	0	2.4E-06	0.00025
6.6E-06	0.00059	1.2E-06	0	0	3.7E-05	5.4E-06	0	0	0	0	3.6E-06	0
0.00018	0.00179	0	6.1E-05	0.00017	2.6E-05	0.0001	6E-07	0.00051	0.00083	0	7.2E-06	0.00039
0.00439	9E-05	8.4E-06	0.00078	0.00016	0.00088	0.00092	0.0006	0	0	0	0.0004	0.00035
0	0.00024	0	0	6E-07	4E-05	0	0	0	4.1E-05	0	1.2E-06	0
1.3E-05	0.00026	0	0	4.2E-06	5.5E-05	6.6E-06	0	0.00037	0.00017	0	4.2E-06	0.00047
0.00378	1E-05	1.4E-05	0.00129	0.0008	0.00186	0.00039	0.00014	0	0	6E-07	0.00024	0.0017
0.00023	0	4.2E-06	0	4.2E-05	7.1E-05	5.1E-05	6E-07	0	0	0	0	0
0.00039	0.0014	1.2E-06	1.2E-06	6.6E-06	0	1.1E-05	0	0	0	0	0	0
2.5E-05	0.00117	4.2E-06	0	6E-07	0.00012	3.6E-06	6E-07	0.00093	0.00348	0	0.00099	0.00171
1.9E-05	0.00271	1.2E-06	0	9.6E-06	1.8E-05	4.2E-06	4.2E-06	0.0009	6.6E-05	0	2.4E-06	0.00038
0.00042	0.00241	1.6E-05	1.2E-06	0.00034	0.00025	1.2E-06	0.00016	0.00283	0.00121	0	0.00015	0.00169
0.00624	5E-05	0.00014	0.00032	8.2E-05	0.00159	0.00123	0.00014	0	0	3.6E-05	0.00079	0.00048
4.8E-05	0.0005	6E-07	0	1.4E-05	3.1E-05	1.3E-05	6E-07	0.00025	3.5E-05	0	3E-06	0.00033
0.00045	0.00225	1.1E-05	9.6E-05	2.8E-05	6.6E-05	0.00021	3E-05	0.0012	0.00086	0	0.00011	0.00087
0.0103	0.0005	7.2E-06	0.00013	5.4E-05	0.00042	5.8E-05	1.8E-06	0.00033	0.00281	1.8E-06	0.00012	0.00295
8.1E-05	0.00115	1.9E-05	1E-05	6.1E-05	0.00027	6E-07	1.3E-05	0.00147	0.0062	2.4E-06	4.6E-05	0.00052
0.00103	6E-07	1.7E-05	0.00029	3.6E-06	0.00086	0.00069	0.00019	0	0	1.5E-05	0.00094	0.00011
1.9E-05	2.5E-05	7.2E-06	0	0	6E-07	0	0	4.8E-06	5.4E-05	0	1.2E-06	1.2E-06
6E-06	0.00014	0	0	0	6E-07	9E-06	0	0	6E-07	0	0	0
0	0.00021	0	0	7.2E-06	0	0	0	0.00015	0	0	0	0
0.00438	0.0009	0	0	0	0	0.00017	0	0	0.00016	0	0	0
6.4E-05	0.00028	1.2E-06	0	2.4E-06	0	1.2E-06	0	0	0.00048	0	3.6E-06	0
0	8.5E-05	0	0	0	0	0	0	0	2.2E-05	0	0	0
0.0007	5.4E-06	0	0.00079	6E-07	0.00017	0.00042	6E-07	0	0	0	0	1.2E-06
7E-05	0	2.4E-06	1.2E-06	0.0001	6E-05	0.0004	1.5E-05	0	0	2.4E-06	0.00039	0.00052
0.00282	0	1.8E-06	0	0	0	6E-07	0	0	0	0	0	0
0.00025	0	1.8E-06	9.6E-06	7.2E-06	3.9E-05	2.5E-05	0.00033	0	0	0	3.4E-05	0
0.00134	6E-07	0	0.00018	3.5E-05	0.00012	4.8E-05	0.0002	0	0	0	9.6E-05	0.00013

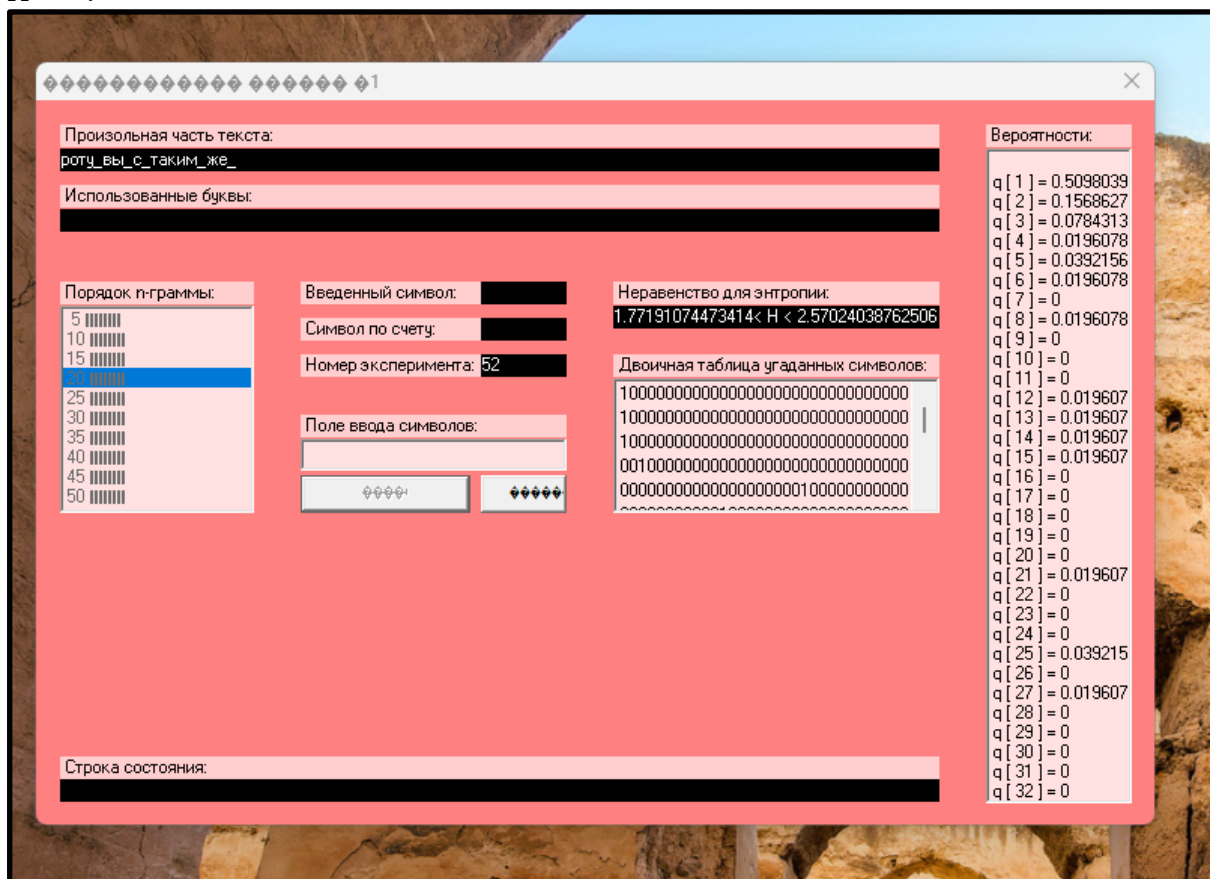
Таблиця значень ентропії та надлишковості для файла > 1 Mb (FILE):

Тип джерела відкритого тексту	Ентропія	Надлишковість
H1 (монограми)	4.4393304946163665	0.10392519462562477
H1_ws (монограми з пробілами)	4.347574526370398	0.1304850947259204
H2_no (біграми без перетинів)	4.122434891829624	0.16789028259001282
H2_no_ws (біграми без перетинів з пробілами)	3.9412994623955098	0.211740107520898
H2_ov (біграми з перетинами)	4.121835034126119	0.16801136331954447
H2_ov_ws (біграми з перетинами та пробілами)	3.9413951409046173	0.21172097181907656

Таблиця значень ентропії та надлишковості для заданого файла TEXT:

Тип джерела відкритого тексту	Ентропія	Надлишковість
H1 (монограми)	4.422101851111231	0.10740278058018504
H1_ws (монограми з пробілами)	4.385354297866487	0.12292914042670267
H2_no (біграми без перетинів)	3.9748832357629778	0.1976734495907414
H2_no_ws (біграми без перетинів з пробілами)	3.8839674721838	0.22320650556324007
H2_ov (біграми з перетинами)	4.02077298377601	0.18841064587082812
H2_ov_ws (біграми з перетинами та пробілами)	3.914743461832841	0.21705130763343183

Після цього нами було проведено серію експериментів з CoolPinkProgram.exe. На основі отриманих значень даних ентропії, була обчислена надлишковість, використовуючи попередньо вказану формулу. Результати представлені у таблиці нижче.

 $H^{(10)}:$  $H^{(20)}:$

Произвольная часть текста:
кую_нибуд

Использованные буквы:

Порядок n-граммы:

Введенный символ:

Символ по счету:

Номер эксперимента: 54

Неравенство для энтропии:
 $1.51081631071507 < H < 2.30339609971048$

Двоичная таблица угаданных символов:

Поле ввода символов:

Вероятности:

Строка состояния:

q[1]	= 0.5283018
q[2]	= 0.1698113
q[3]	= 0.0943396
q[4]	= 0.0377358
q[5]	= 0.0566037
q[6]	= 0.0188679
q[7]	= 0
q[8]	= 0
q[9]	= 0
q[10]	= 0.018867
q[11]	= 0
q[12]	= 0
q[13]	= 0.018867
q[14]	= 0
q[15]	= 0
q[16]	= 0.018867
q[17]	= 0.018867
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]	= 0.018867

$H^{(30)}$:

Произвольная часть текста:
м_не_имеет_под_собой_твердого

Использованные буквы:

Порядок n-граммы:

Введенный символ:

Символ по счету:

Номер эксперимента: 52

Неравенство для энтропии:
 $1.36810044834647 < H < 2.00910114755131$

Двоичная таблица угаданных символов:

Поле ввода символов:

Вероятности:

Строка состояния:

q[1]	= 0.6470588
q[2]	= 0.0588235
q[3]	= 0.0980392
q[4]	= 0.0392156
q[5]	= 0.0196078
q[6]	= 0
q[7]	= 0
q[8]	= 0
q[9]	= 0.0196078
q[10]	= 0
q[11]	= 0.019607
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]	= 0.019607
q[24]	= 0.019607
q[25]	= 0
q[26]	= 0.039215
q[27]	= 0
q[28]	= 0
q[29]	= 0
q[30]	= 0
q[31]	= 0.019607
q[32]	= 0

	Ентропія	Надлишковість
$H^{(10)}$	$1.51081631971597 < H < 2.30339609971048$	$0,53932078 < R < 0,697836736$
$H^{(20)}$	$1.77191074473414 < H < 2.57024038762506$	$0,485951922 < R < 0,645617851$
$H^{(30)}$	$1.36810044834647 < H < 2.00910114755131$	$0,59817977 < R < 0,72637991$

<https://gvern.net/doc/cs/algorithm/information/compression/1994-levitin.pdf>

Entropy of Natural Languages: Theory and Experiment Lev. B. Levitin.

За наведеним згори ресурсом ми знайшли таблиці з оцінками ентропії та надлишковості різних мов. Порівнявши з отриманими в якості експерименту даними ми бачимо, що значення ентропії та надлишковості російської мови наближаються до істинних пропорційно зі збільшенням кількості символів у n-грамах.

Table 1. L-gram entropies for eight languages								
F_L	English ¹	Russian ²	French ³	Rumanian ⁴	Hebrew ⁵	Arabic ⁶	Portugese ⁷	German ⁸
F_0	4.76	5.00	4.76	4.76	4.52 ⁹	5.00	4.76	4.7
F_1	4.03	4.35	3.9		4.12 ¹⁰	4.2	3.9	4.1
					4.22 ¹⁰			
F_2	3.32	3.52			3.71 ¹⁰	3.8	3.5	
F_3	3.1	3.0	2.83	2.69	3.21 ¹⁰	2.5	3.1	2.95

Entropy of natural languages						721
Table 3. Entropy and redundancy of fiction texts for several languages by Piotrovski <i>et al.</i>						
	Language	\hat{H} bits	\underline{H} bits	\hat{R} (%)	\hat{R} (%)	
1	Russian	1.19	0.70	76	86	
2	Polish	1.29	0.83	74	84	
3	English	1.10	0.65	77	86	
4	German	1.36	0.83	71	82	
5	French	1.36	0.78	71	84	
6	Rumanian	1.26	0.78	74	84	
7	Kazakh	1.35	0.81	75	85	

Висновки

Під час виконання нами даного комп'ютерного практикуму

першочергово відбулося наше знайомство з необхідною теоретичною базою. Зокрема були розглянуті поняття ймовірнісного ансамблю, ентропії, в тому числі сукупної, умовної та питомої, а також поняття надлишковості джерела відкритого тексту (мови).

Отримані теоретичні знання були закріплені на практиці за рахунок здійснення дослідження різних текстів російською мовою, що було зокрема виконано з використанням власноруч створеного програмного засобу для автоматизації аналізу. Порівнюючи отримані результати з відкритими даними, зокрема з роботи Лева Левітіна "Entropy of Natural Languages: Theory and Experiment", можна відзначити наближеність отриманих ентропій, водночас надлишковість обчислена з використанням даних значень, відповідно є віддаленою від загально прийнятих значень та від тієї, що була обчислена з використанням наданої нам CoolPinkProgram.exe, що пояснюється використанням ентропій n-грам малих розмірностей (монограм і біграм).