## КРИПТОГРАФІЯ КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

## Експериментальна оцінка ентропії на символ джерела відкритого тексту

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.

Виконали Шанідзе Давид та Тивонюк Володимир

Текст, який був використаний для роботи: "Над пропастью во ржи" Джерома Дэвида Сэлинджера

### Хід 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 Н1 та Н2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення Н1 та Н2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення Н1 та Н2 на тому ж тексті, в якому вилучено всі пробіли.
- 2. За допомогою програми CoolPinkProgram оцінити значення (10) Н, (20) Н, (30) Н.
- 3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

## Опис труднощів, що виникали, та шляхів їх розв'язання

Перш за все треба було знайти текст, який би ми аналізували. Знайти було не так важко, як скопіювати його весь у текстовий файл. Після цього цікавим моментом було редагування файлу з текстом, тобто видалення усіх зайвих символів та подвійних пробілів. Та складніше за все було впоратися з другим, але нам все-таки вдалося знайти досить хитрий вихід з даної ситуації (можете побачити у файлі edit.py). Далі прийшла черга виконання самої суті завдання. Якщо перше завдання було виконано досить просто та друге також не викликало запитань, то з третім вже довелося розбиратися. Після декількох годин гуглення та перечитування методички ми зрозуміли, що відповідь лежала майже перед нашими очима. Тоді ми закінчили з цим завданням й перейшли до третього. Там вже треба було працювати з даним нам екзе файлом. Найскладніше тут було вгадувати початок слова (тобто, що йшло після пробілу) та й взагалі робота тут була досить нудна й однообразна. У четвертому і вже останньому завданні проблемою було визначити  $H_{\infty}$  та  $H_{0}$ . Але потім ми розібралися з даною проблемою, результати чого можна бачити далі.

# Таблиці частот букв і біграм тексту, одержані значення Н1 та Н2

## Таблиця Частоти Біграм:

(intersection, spaces)

\ a	6	В	г	Д	e	×	3	И	Й	K	л	M	H	0	п	р	C	T	У	Φ	×	ц	4		Щ	ы	ь	э	ю	8
a 2.9e-8	5 0.0006337	0.0049223	0.0008615	0.0052102	2.9e-05					0.0075212	0.0049782 (	0.0031559			0.0011597 (	.0073762	0.0015076	0.006581	2.9e-05	0.000107	0.0004224	0.0004887	0.0015345	0.0009795			4.1e-06	9	4.1e-06	9
6 0.000403	8 7.87e-05	1.66e-05	9	5.38e-05 (	0.0014268						1.45e-05			0.0030586		6.42e-05	7.87e-05	6.21e-05	0.0005198		9	6	9	9		0.0001657	1.66e-05	9	0.0003313	
	4 4.1e-06	2.9e-05		0.0005736			0.0006275		9 6	0.0001015			1.24e-05				0.0010458				0.0001015	1.24e-05	. 0	3.31e-05		0.0009712				0.0001242
г 0.000397		9		4.1e-06			0.0003562		0		4.97e-05		0.0002009				1.66e-05				9	6	9	9		0 9.94e-05			3.31e-05	
	9 2.07e-05										0.0002816		0.0003334				0.000234				9	8.3e-06		0		0 8.08e-05		0.0003665		
									1.24e-05 6											0.000109	4.14e-05	0.000615	0.0041271	0.0020045		5 0.0006896	0.0005384	0		8.9e-05
	1 8.3e-06		0			1.24e-05			0		0.0001595		4.1e-06				1.24e-05		0.0013522		9	6	9	0		0 2.9e-05	0		1.24e-05	
	1 2.07e-05		0		0.0009049		8.9e-05			3.73e-05			2.9e-05			.0001367			0.0002506		9	6	9	0		0 3.31e-05		1.24e-05		
	5 0.0009981	0.002574	0.0004784			0.0009339			0	0.00304	0.0070532 (	0.0018037			0.0010023 (	.0044916	0.0019859			0.0002444	0.0002154	0.0001512	0.0017229	0.0012549				0	4.1e-06	
й 0.00079		0	9		0.0025222	0		0.0006958	0	0	0	9		0.0031435		9	9		0.0001491		9	6	9	9		0.0014848			2.48e-05	
															0.0001532 (								0.0002754			0.0001429				
	6 0.0004369														0.0006896								4.1e-06			0.0028991			1.66e-05	
	9 1.66e-05					4.1e-06					2.1e-06			0.0052454					0.0014019		8.3e-06		2.1e-06			0.0004804			4.97e-05 (	
	3 0.0001305														4.76e-05 (						0.0001408		0.0009319			5 0.0001615				0.0004349
	5 0.0016525										0.0057506				0.008188					0.000184	0.0026941	0.0001719					8.3e-06	9	0.0001325	9
n 0.00084			1.24e-05					0.0004059							2.48e-05					•	9	6	9	1.66e-05		0.000147		9		4.14e-05
	7 0.0008656						0.0001636			0.001551					0.0052578					8.28e-0	6.83e-05	6	9	9		0.000234			2.28e-05	
c 0.003621															0.0001035						9	6	9	9		0.0003603			0.0003562 (	
T 0.005032			6.63e-05						0.0007289 6												4.76e-05		0.0041665			0.0004784	0.0001284	0.0024891	0.00076	0.0018078
	5 0.0012756	0.0004452				0.0005115			9 6						0.0005467						0.000147	0.0002257	0.0008511	0.0002154	3.73e-0		0	0	0	0
ф 8.7e-0		0	4.1e-06			0		0.0001139	0		6.63e-05		8.3e-06				1.24e-05		2.48e-05	8.3e-0	. 0	6	9	0			7.04e-05	9	0	0
	2 2.28e-05			3.52e-05 (		0		0.0014082		4.1e-06	0		2.48e-05								9	6	9	0		0.0005943	0	9	2.48e-05	
ц 0.000188		1.24e-05		0.0002174					4.14e-05		9		0.0004307				2.28e-05			•	9	6	9	9		0 4.1e-06		9		2.07e-05
4 0.000757		0.0002485		1.24e-05 (				0.0010395			0.0001284								0.0009153				4.1e-06	9		0 8.28e-05			9.11e-05 (	
w 0.001207		3.93e-05	9	6.63e-05		9		0.0008635	2.07e-05	9	2.9e-05						7.45e-05		0.0009339	1.66e-0	4.1e-06	6	0.0001325	9		0.0004349			2.07e-05	
	7 0.0004991		8		0.0006875	9		9.94e-05	9	9	9		5.8e-05			1.04e-05	9		6.42e-05		9	6	9	9	•	0 5.18e-05	4.1e-06	9	8.7e-05 (	0.0002071
	0.004504			0.0003583	9		0.0004369		9		0.0009112 (				0.0001035				9	1.24e-0		0.0001367		9	•	3 0	9	9	0	9
ь	0 0.0001015			0.0008014	0	1.66e-05 (			0	0	0.003274	8.49e-05			0.0001553 (				9		9	6	0.0003106	0.0011835	2.48e-0	9 د	0	0	0	9
9		8.3e-06		8.3e-06	0	0		1.66e-05	0	0	0	9		1.24e-05					8.3e-06		9	6	9	9		3 0	0	9	0	0
	6 8.3e-06			2.48e-05 (		0		0.000205	0						4.1e-06						9	6	9	4.1e-06			0.0004908	0	1.66e-05 (	
я 0.001406	1 0.0006999	0.0001698	0 (	0.0001739	5.38e-05	0 (	0.0003438	0.0005695	_0	9	0.0008884 (	0.0004183	0.002135	0.0006709	0.0004307	0.001023	0.0043342	0.000352	8.3e-06		9	6	9	0		0 1.24e-05	0.0002526	9	0	1.04e-05

## (intersection, NO spaces)

\ a 6 B	r A	e	×	3	И	й	K	л	М	н	0	п	р	c	T	У	•	x	ц	ч		щ	ы	ь	3	10	я
a 4.51e-05 0.0009855 0.007655	2 0.0013397 0.008102	8 4.51e-05	0.0022157	0.0062929 (	0.0001578	1.29e-05 0	.0116969	0.0077421	0.0049081	0.0166468	9.02e-05	0.0018035	0.0114715	0.0023445	0.0102348	4.51e-05	0.0001675	0.000657	0.00076	0.0023864 (	0.0015233	0.0004026	9	6.4e-06	9	6.4e-06	9
6 0.000628 0.0001224 2.58e-0	5 0 8.37e-0	5 0.0022189	1.29e-05	0.0002931 (	0.0014235	8.7e-05	9	2.25e-05	9.34e-05	9	0.0047567		9.98e-05	0.0001224	9.66e-05	0.0008083	9	9	0	9	0	0	0.0002576	2.58e-05	9	0.0005153	1.93e-05
B 0.0048823 6.4e-06 4.51e-0	9 0.000892	1 0.0020321	9	0.0009758 (	0.0029693	0 0	.0001578	4.51e-05	0	1.93e-05	0.0102767		0.000409	0.0016264	0.0018325	0.0004638	9	0.0001578	1.93e-05	9	5.15e-05	0	0.0015104	3.86e-05	3.86e-05	0	0.0001932
г 0.0006183 0	0 6.4e-0	6 0.0046987	9	0.0005539 (	0.0009694	0	0	7.73e-05	0	0.0003124	0.0073041	. 0	9.02e-05	2.58e-05	6.4e-06	0.0016586	9	9	0	9	0	0	0.0001546	0.0001803	9	5.15e-05	3.86e-05
д 0.00314 3.22e-05 0.000998	4 0.0036134 7.09e-0	5 0.0024991	0.0006699	0.0009501 (	0.0029532	0.0003478	9	0.000438	9	0.0005185	0.006032	. 0	0.0003414	0.0003639	0.0001449	0.0034524	9	9	1.29e-05	9	9	0	9.0001256	0.0001642	0.00057	0.0003768	0.0005958
e 0.0035232 0.0026247 0.005764	7 0.0003156 0.007674	5 0.002322	0.0049789	0.0003575 (	0.0019645	1.93e-05 0	.0010145	0.004995	0.0054072	0.0176355	0.002892	0.0024766	0.0064636	0.0072075	0.0079225	0.0003349	0.0001707	6.44e-05	0.0009565	0.0064185 (	0.0031175	0.0020225	9.0010724	0.0008373	9	9	0.0001385
x 0.0018679 1.29e-05	0 0.000154	6 0.0005668	1.93e-05	0.0001803 (	0.0007021	9	9	0.000248	9	6.4e-06	0.0027213	9	0.0002673	1.93e-05	9	0.002103	9	9	9	9	9	9	4.51e-05	9	9	1.93e-05	4.83e-05
3 0.0055522 3.22e-05 0.000405	8 0	0 0.0014074	9	0.0001385	0.0018131	9	5.8e-05	1.93e-05	9	4.51e-05	0.001475	9	0.0002126	1.93e-05	9	0.0003897	9	9	9	9	9	9	5.15e-05	0.0001803	1.93e-05	1.29e-05	0.0001868
и 7.73е-05 0.0015523 0.004003	1 0.0007439 0.003600	5 5.15e-05	0.0014525	0.0004831 (	0.0003382	9 9	.0047277	0.0109691	0.0028051	0.0101543	0.0010467	0.0015587	0.0069853	0.0030885	0.0047921	1.93e-05	0.00038	0.0003349	0.0002351	0.0026795	0.0019516	0.0006763	6.4e-06	3.22e-05	9	6.4e-06	3.86e-05
й 0.0012302 0	9 0	0 0.0039226	8	9 6	0.0010821	9	0	8	9	9	0.0048887		9	9	9	0.0002319	9	8	9	9	9	9	0.0023091	9	3.22e-05	3.86e-05	1.29e-05
K 0.0093556 0.0001256 0.000199	7 7.41e-05 0.000367	1 0.0012174	0.0004026	7.73e-05 (	0.0029951	0.0001481	5.47e-05	0.0004734	0.0001965	0.0009533	0.0026408	0.0002383	0.000715	0.004615	0.0013269	0.0009758	9	8	0.0002158	0.0004283	0.0017101	9	9.0002222	0.0021159	5.15e-05	1.29e-05	0.000541
л 0.0138965 0.0006795 0.000505	6 0.0010209 0.001049	9 0.0089981	3.86e-05	0.0002641 (	0.0081672	0.0001159 0	.0014847	0.0011787	7.73e-05	9	0.007755	0.0010724	0.0001159	0.0041995	0.0002963	0.0014363	2.58e-05	9.02e-05	9	6.4e-06 (	0.0009243	9	0.0045087	9	5.15e-05	2.58e-05	0.0015008
M 0.0048662 2.58e-05 0.000115	9 8.05e-0	5 0.0051625	6.4e-06	0.0001288 (	0.0028791	0.0001224	9	3.2e-06	2.58e-05	9	0.0081576	9	0.000438	0.0013558	3.86e-05	0.0021803	6.4e-06	1.29e-05	9	3.2e-06	1.29e-05	9	0.0007472	0.0003446	1.29e-05	7.73e-05	0.0004895
н 0.003607 0.0002029 0.001239	9 0.0002158 0.002470	1 0.0091495	0.0007697	0.0026666 (	0.0044636	0.0010531 0	.0006538	0.000248	0.0037648	0.0020482	0.0114006	7.41e-05	0.0017487	0.0018389	0.001446	0.0002834	9	0.000219	0	0.0014492 (	0.0007182	1.93e-05	0.0002512	0.0013269	0.0002576	6.44e-05	0.0006763
0 6.44e-05 0.00257 0.010102	8 0.0116293 0.005001	5 0.0001675	0.0001997	0.0004605	0.000409	9 9	.0123313	0.0089434	0.0055039	0.0146888	0.0006248	0.0127339	0.0096905	0.0030981	0.0226112	1.61e-05	0.0002866	0.0041899	0.0002673	0.0004187 (	0.0003897	6.4e-06	9	1.29e-05	9	0.0002061	0
n 0.0013204 0 0.000119	2 1.93e-05 3.54e-0	5 0.0008695	9	9.7e-06 6	0.0006312	2.58e-05	9	1.93e-05	0.000132	1.29e-05	0.0014041	3.86e-05	0.0001095	0.0026634	0.0001353	0.0006248	9	9	0	9	2.58e-05	0	0.0002287	1.29e-05	9	9	6.44e-05
p 0.0032173 0.0013462 0.000888	9 0.0011497 0.001954	9 0.0093942	9	0.0002544 (	0.0007439	9 9	.0024122	9	5.8e-05	0.0005668	0.0083927	0.0081769	7.41e-05	0.0005346	0.0042801	0.0011916	0.0001288	0.0001063	9	9	9	0	0.0003639	9	0.0001514	3.54e-05	6.4e-06
c 0.0056327 4.51e-05 0.005732	5 1.93e-05 0.000341	4 0.0056423	1.29e-05	4.51e-05 (	0.0029919	0.0006087 0	.0003027	0.002235	0.0001288	0.0010112	0.0075779	0.000161	0.0002222	0.0009339	0.0024959	0.0015426	9	9	9	9	9	0	0.0005604	0.0014041	1.29e-05	0.0005539	0.0002061
т 0.0078258 0 0.000486	3 0.0001031 0.000363	9 0.0106245	1.29e-05	3.86e-05 (	0.0072204	0.0011336 0	.0012367	0.000132	9.02e-05	0.0006763	0.0095617	9.02e-05	0.0019033	0.0125117	0.0004187	0.003053	0.0001353	7.41e-05	9	0.0064797	0.0001868	9	0.0007439	0.0001997	0.0038711	0.0011819	0.0028115
v 0.0001481 0.0019838 0.000692	4 0.0008277 0.002975	8 8.05e-05	0.0007955	0.0007665	9	9 9	.0033976	0.0018325	0.0028115	0.0026634	0.0001868	0.0008502	0.0035877	0.0006763	0.0025249	1.29e-05	0.0001031	0.0002287	0.000351	0.0013236	0.0003349	5.8e-05	9	9	9	9	9
φ 0.0001353 0	0 6.4e-06 6.4e-0	6 0.0002319	9	9 6	0.0001771	9	9	0.0001031	9	1.29e-05	0.0002126	9	3.22e-05	1.93e-05	9	3.86e-05	1.29e-05	9	9	9	9	9	9	0.0001095	9	9	9
x 0.0009178 3.54e-05 1.93e-0	5 0 5.47e-0	5 0.0006409	8	9 6	0.0021899	9	6.4e-06	8	9	3.86e-05	0.0008309	9	0.0001127	0.0001578	1.29e-05	0.0003221	9	8	9	9	9	9	0.0009243	9	9	3.86e-05	0.0001095
ц 0.0002931 6.4е-06 1.93е-0	5 0 0.000338	2 0.0005282	6.4e-06	9 6	0.0009629	6.44e-05	6.4e-06	8	9	0.0006699	9.66e-05	9	6.44e-05	3.54e-05	9.02e-05	3.86e-05	9	8	9	9	9	9	6.4e-06	5.47e-05	9	9	3.22e-05
4 0.0011787 1.29e-05 0.000386	5 6.4e-06 1.93e-0	5 0.0015652	6.4e-06	1.29e-05 (	0.0016167	0.0002126	0	0.0001997	3.22e-05	0.0002641	0.0033719	6.4e-06	0.0002093	0.0004122	0.000161	0.0014235	1.93e-05	8	9	6.4e-06	9	9	0.0001288	0.0002802	9	0.0001417	0.0002222
ш 0.0018776 0 6.12e-0	9 0.000103	1 0.0020354	9	6.4e-06	0.001343	3.22e-05	0	4.51e-05	6.4e-06	1.29e-05	0.0019774	1.29e-05	0.0003478	0.0001159	9	0.0014525	2.58e-05	6.4e-06	0	0.0002061	9	9	0.0006763	0.0006763	9	3.22e-05	6.4e-06
щ 0.0002126 0.0007761 1.93е-0	5 0	0 0.0010692	9	0 (	0.0001546	9	9	9	9	9.02e-05	0.0001578	9	1.61e-05	0	9	9.98e-05	9	9	0	9	9	9	8.05e-05	6.4e-06	9	0.0001353	0.0003221
ы 0 0.0070046 0.003104	6 0.000557	1 0	e	0.0006795	9	9	9	0.001417	0.0014589	0.002644	9	0.000161	0.0014106	0.0006022	0.0024057	0	1.93e-05	9	0.0002126	9	9	9	9	0	9	0	9
ь 0 0.0001578 0.0001	9 0.001246	3 0	2.58e-05	0.0002061	9	9	9	0.0050916	0.000132	0.0023059	9	0.0002415	0.0004831	0.0037422	0.0112074	9	9	9	0	0.0004831 (	0.0018405	3.86e-05	9	9	9	0	0
9 0 0 1,29e-0	5 0 1.29e-0	5 0	9	9	2.58e-05	9	9	9	9	9	1.93e-05	9	0.0005571	0.0001224	6.4e-06	1.29e-05	9	9	9	9	0	9	9	9	9	0	0
m 0.0026666 1.29e-05 1.29e-0	9 3.86e-8	5 0.0001578	9	9 (	0.0003188	0	0	0.0012399	1.29e-05	0.0004219	0.0004348	6.4e-06	0.0011433	0.0003736	4.19e-05	0.001256	0	9	9	0	6.4e-06	9	0	0.0007633	9	2.58e-05	0.0002029
я 0.0021867 0.0010885 0.000264	0 0.000270	5 8.37e-05	9	0.0005346 (	0.0008856	0	0	0.0013816	0.0006505	0.0033203	0.0010434	0.0006699	0.0015909	0.0067405	0.0005475	1.29e-05	0	9	9	0	0	9	1.93e-05	0.0003929	9	0	1.61e-05

## (NO intersection, spaces)

\ a	6	В	г	д	e	×	3	и	й	K	л	м	н	0	п	р	c	T	У	φ	x	ц	ч		Щ.	ы	ь	э	10	я
a 3.31e-05	0.0007165 6	0.0049451	0.0007703 (	0.0054214	8.3e-06	0.0015075	0.0039884	9.94e-05	0	0.0071691	0.004883 (	0.0030607	0.010586	8.28e-05	0.0011182	0.0073928	0.0015241	0.0063988	2.48e-05	0.0001077	0.0004017	0.0005426	0.0014082	0.0009898	0.0002651	0	9	0	8.3e-06	9
6 0.0003893	7.45e-05	3.31e-05	0	6.63e-05	0.0014703	0	0.0001864	0.0008946	5.38e-05	0	1.66e-05	2.9e-05	9	0.0031435	0	7.04e-05	7.45e-05	5.8e-05	0.000584	9	0	9	9	0	0	0.0001822	2.48e-05	0 (	.0003065	1.66e-05
B 0.0031352	8.3e-06	8.3e-06	0 (	0.0006212	0.0014082	0	0.0005757	0.0019507	0	0.0001367	1.66e-05	9	1.66e-05	0.0064816	0	0.0002609	0.0009857	0.0012094	0.0003313	9	7.87e-05	2.48e-05	9	4.14e-05	0	0.001052	3.31e-05	4.14e-05	0 0	.0001491
r 0.0004639	9	9	9	9	0.002866	9	0.0003148	0.0006171	9	9	4.97e-05	9	0.0001905	0.0047049	9	3.31e-05	1.66e-05	8.3e-06	0.0009774	9	9	9	9	9	9	0.0001201	9.94e-05	9	3.31e-05	8.3e-06
д 0.0020584	3.31e-05 6	0.0006171	0.0023027	3.31e-05	0.0015862	0.0004224	0.0006088	0.0020253	0.0002361	9	0.000323	9	0.000352	0.0037192	9	0.0002526	0.0002236	0.000116	0.0022448	9	9	1.66e-05	9	9	9	7.87e-05	9.53e-05	0.0003023 (	.0002775 0	.0002733
e 0.0023483	0.0016608 6	9.0036281	0.0001905	0.0049492	0.0015862	0.0032222	0.0002444	0.0011638	9	0.0006005	0.0032802 (	9.0035866	0.0116918	0.0019093	0.0016111	0.00427	0.004651	0.0050445	0.0001781	0.000116	2.48e-05	0.0005923	0.0043818	0.0019466	0.0013419	0.0006378	0.0005301	9	0 9	9.11e-05
x 0.0012591	8.3e-06	8	9	9.94e-05	0.0004183	1.66e-05	7.45e-05	0.0004307	8	9	0.0001077	9	8.3e-06	0.0017063	9	0.000145	9	9	0.0013957	9	9	9	9	9	9	3.31e-05	9	9	2.48e-05	3.73e-05
3 0.003715	1.66e-05 6	0.0002485	9	9	0.0009691	8	0.0001077	0.0012383	8	3.31e-05	1.66e-05	9	2.48e-05	0.0009816	9	0.0001491	2.48e-05	9	0.0001657	8	9	9	8	9	9	2.48e-05	9.53e-05	1.66e-05	1.66e-05 S	9.94e-05
и 4.14е-05	0.0009567 6	0.0024601	0.0005053	0.0023234	3.31e-05	0.0008946	0.0002733	0.0002278	8	0.0032098	0.0071733 (	0.0017933	0.0066846	0.0007786	0.0009153	0.0044647	0.001843	0.0031393	9	0.0002319	0.0002319	0.0001698	0.0015365	0.0012756	0.0004514	9	2.48e-05	0	8.3e-06	2.48e-05
й 0.0006461	9	9	9	9	0.0026051	9	9	0.0006502	9	9	0	9	9	0.0032884	9	9	9	9	0.0001077	9	9	9	9	9	9	0.0014786	9	1.66e-05	3.31e-05	8.3e-06
K 0.0060136	8.28e-05 6	0.0001118	5.38e-05 (	0.0002319	0.0007703	0.0002982	5.8e-05	0.00193	0.0001201	4.56e-05	0.0003479 (	0.0001408	0.0005674	0.0016898	0.0001367	0.0004804	0.0030234	0.0006917	0.0006378	9	9	0.000145	0.0003065	0.0009981	9	0.0001325	0.0012922	4.14e-05	0 0	.0002816
л 0.0090163	0.0004514 6	0.0003189	0.0005674	0.0007496	0.0055581	8.3e-06	0.0001739	0.0052226	7.45e-05	0.0010188	0.0008035	4.97e-05	9	0.0048788	0.0007703	7.45e-05	0.0026879	0.0001988	0.0009484	8.3e-06	6.63e-05	9	8.3e-06	0.0006502	9	0.0030234	9	3.31e-05	2.48e-05 0	.0009153
M 0.0031849	1.66e-05	9.94e-05	0	7.45e-05	0.0032387	8.3e-06	9.53e-05	0.0018099	8.28e-05	0	0	3.31e-05	9	0.0052143	0	0.0003065	0.0008532	3.31e-05	0.0012798	9	8.3e-06	9	9	1.66e-05	9	0.0004307	0.0002485	8.3e-06	4.97e-05 0	.0003438
н 0.0024187	0.0001242	0.00082	0.0001367 (	0.0016566	0.0057279	0.0003976	0.0017105	0.0030399	0.0007331	0.0004349	0.0001698 (	0.0023731	0.0011886	0.0075336	4.56e-05	0.0010354	0.0012632	0.0010271	0.000145	9	0.0001657	9	0.0009526	0.0003727	1.66e-05	0.0001408	0.0008366	0.0001698	3.31e-05 0	.0003976
o 4.97e-05	0.0015531 6	0.0066887	0.0077821	0.0032098	9.11e-05	9.11e-05	0.0003065	0.0003065	0	0.0078691	0.0056823 (	0.0036073	0.0093849	0.0004059	0.0077614	0.0061255	0.0020377	0.014388	1.66e-05	0.0001739	0.0027003	0.0001532	0.0002568	0.0002816	0	0	8.3e-06	9 6	.0001242	0
п 0.0009112	9	5.8e-05	1.66e-05	2.9e-05	0.000497	9	8.3e-06	0.0004224	1.66e-05	0	8.3e-06 (	0.0001077	0	0.000965	2.48e-05	7.45e-05	0.0017271	7.04e-05	0.0004017	9	0	9	9	2.48e-05	9	0.0001325	8.3e-06	0	0	3.31e-05
p 0.0021164	0.0007579 6	0.0005177	0.000791	0.0011141	0.0059432	0	0.0001242	0.0005218	0	0.0016194	0	4.14e-05	0.0003396	0.0052764	0.0052681	2.9e-05	0.0002692	0.0028412	0.0007413	7.45e-05	5.8e-05	9	9	0	9	0.0002485	0	7.04e-05	2.07e-05	0
c 0.0035866	2.48e-05 6	0.0035866	2.48e-05	0.0002154	0.0038227	8.3e-06	4.97e-05	0.0019838	0.0004142	0.0002071	0.0014661 (	0.0001077	0.0006047	0.0048001	9.11e-05	0.0001905	0.0005964	0.0017063	0.0010106	8	0	0	8	0	0	0.0003852	0.0009194	1.66e-05 (	.0003479 0	.0001077
T 0.0049989	0	0.000323	6.63e-05	0.0001615	0.0067301	8	2.48e-05	0.0046345	0.0007165	0.0007952	5.8e-05	4.14e-05	0.0003852	0.0064071	5.8e-05	0.0013129	0.0076827	0.0002485	0.0021122	0.0001325	5.8e-05	9	0.0045268	0.0001325	0	0.0004307	0.0001408	0.0023442 (	.0007745 0	.0018803
y 9.94e-05	0.0013129 6	0.0004266	0.000584	0.0017933	1.24e-05	0.0005633	0.0004846	0	8	0.0023773	0.0011762 (	0.0017643	0.0017229	0.000116	0.0005384	0.0023814	0.000555	0.0016815	9	5.8e-05	0.0001284	0.0002692	0.0007828	0.0002029	3.31e-05	0	8	9	0	0
ф 6.63e-05	9	9	8.3e-06	8.3e-06	0.0001325	9	9	8.7e-05	9	0	8.28e-05	9	0	0.000116	0	8.3e-06	1.66e-05	9	4.97e-05	8.3e-06	0	9	9	8	0	9	6.63e-05	9	0	0
x 0.0006047	8.3e-06	8.3e-06	0	6.21e-05	0.0003686	9	9	0.0012591	9	9	9	9	1.66e-05	0.000526	0	0.0001077	9.53e-05	1.66e-05	0.0002029	9	0	9	9	0	0	0.0005508	0	9	3.31e-05 (	6.63e-05
ц 0.0001739	8.3e-06	1.66e-05	0 (	0.0001905	0.0003438	8.3e-06	9	0.0006627	3.31e-05	8.3e-06	9	9	0.000381	7.45e-05	0	4.14e-05	8.3e-06	4.14e-05	3.31e-05	9	0	9	9	0	0	8.3e-06	2.07e-05	9	0	2.48e-05
4 0.0007703	8.3e-06 6	9.0002568	8.3e-06	1.66e-05	0.0009526	8.3e-06	1.66e-05	0.0011389	0.0001367	0	0.0001491	8.3e-06	0.0002154	0.0021122	0	0.0001574	0.0002651	8.28e-05	0.0009609	8.3e-06	0	9	9	0	0	9.11e-05	0.0002112	9	7.87e-05 0	.0001615
<b>■ 0.0012176</b>	9	4.14e-05	0	2.48e-05	0.0013088	9	0	0.0008904	1.66e-05	0	3.31e-05	9	9	0.0012342	1.66e-05	0.0002319	0.000116	9	0.0008946	8.3e-06	0	9	0.0001035	0	9	0.0004142	0.0004887	0	2.48e-05	8.3e-06
щ 0.000116	0.0004639	8.3e-06	0	0	0.0006668	0	0	8.7e-05	9	0	0	9	7.45e-05	0.0001077	0	1.24e-05	0	0	5.38e-05	9	0	9	9	0	0	3.31e-05	8.3e-06	0	8.28e-05 0	.0002402
ы 0	0.0043735 6	0.0020998	0	0.000352	0	0	0.000468	0	9	0	0.0010561 (	0.0009277	0.001814	0	9.11e-05	0.0008283	0.0003396	0.0014206	9	9	0	0.0001325	9	0	0	0	0	0	0	0
ь 0	7.87e-05 6	0.0001491	0 (	0.0008118	0	1.66e-05	0.0001077	9	9	0	0.0034127	5.8e-05	0.0015862	9	0.0001739	0.0003023	0.00234	0.0073431	9	9	0	9	0.0002568	0.0012259	3.31e-05	9	9	0	0	0
9 0	0	1.66e-05	0	0	0	0	0	1.66e-05	0	0	0	9	0	1.66e-05	0	0.0004142	9.94e-05	9	8.3e-06	9	0	9	9	0	0	0	0	0	0	0
m 0.0016484	0	8	9	1.66e-05	8.7e-05	8	0	0.0001491	8	0	0.0007869	8.3e-06	0.0002485	0.0002858	9	0.0007413	0.0002112	2.48e-05	0.0008656	9	9	9	9	9	9	9	0.000526	0	2.07e-05 0	.0001367
я 0.0014289	0.0006461 6	0.0001988	0	0.0002236	7.04e-05	8	0.0003893	0.0006171	9	0	0.0008863 (	0.0004183	0.0021454	0.0006171	0.0004017	0.0010478	0.0042369	0.0003106	8.3e-06	9	9	9	9	9	9	1.66e-05	0.0002899	9	9	8.3e-06

(NO intersection, NO spaces)

a 4.51e-05 0.0009855 0.0076552 0.00133									0.000657 0.000	76 0.0023864 0.0015233	0.0004026 0 6.4e-06	0 6.4e-06 0
6 0.000628 0.0001224 2.58e-05	0 8.37e-05 0.0022189	9 1.29e-05 0.0002931 0.0014235	8.7e-05 0 2	.25e-05 9.34e-05	0 0.0047567	0 9.98e-05	0.0001224 9.66e-05	0.0008084 6	0	0 0 0	0 0.0002576 2.58e-05	0 0.0005153 1.93e-05
B 0.0048823 6.4e-06 4.51e-05	0 0.0008921 0.0020321	0 0.0009758 0.0029693	0 0.0001578 4.	.51e-05 0	1.93e-05 0.0102767	0 0.000409	0.0016264 0.0018325	0.0004638 6	0.0001578 1.93e-	05 0 5.15e-05	0 0.0015104 3.86e-05	3.86e-05 0 0.0001932
r 0.0006183 0 0	0 6.4e-06 0.0046987	7 0 0.0005539 0.0009694	0 0 7.	.73e-05 0	0.0003124 0.0073041	0 9.02e-05	2.58e-05 6.4e-06	0.0016586	0	0 0 0	0 0.0001546 0.0001803	0 5.15e-05 3.86e-05
д 0.00314 3.22e-05 0.0009984 0.00361	34 7.09e-05 0.0024991	1 0.0006699 0.0009501 0.0029532	0.0003478 0 0.	.000438 0	0.0005185 0.006032	0 0.0003414	0.0003639 0.0001449	0.0034524 6	0 1.29e	-05 0 0	0 0.0001256 0.0001642	0.00057 0.0003768 0.0005958
e 0.0035232 0.0026247 0.0057647 0.00031	56 0.0076745 0.002322	2 0.0049789 0.0003575 0.0019645	1.93e-05 0.0010145 0.	.004995 0.0054073	0.0176356 0.002892	0.0024766 0.0064636	0.0072075 0.0079225	0.0003349 0.0001707	6.44e-05 0.00099	65 0.0064185 0.0031175	0.0020225 0.0010724 0.0008373	0 0.0001385
* 0.0018679 1.29e-05 0	0 0.0001546 0.0005668	3 1.93e-05 0.0001803 0.0007021	0 0 0.	.000248 0	6.4e-06 0.0027213	0 0.0002673	1.93e-05 0	0.002103	0	0 0 0	0 4.51e-05 0	0 1.93e-05 4.83e-05
3 0.0055522 3.22e-05 0.0004058	0 0.0014074	0 0.0001385 0.0018132	0 5.8e-05 1	.93e-05 0	4.51e-05 0.001475	0 0.0002126	1.93e-05 0	0.0003897	0	0 0 0	0 5.15e-05 0.0001803	1.93e-05 1.29e-05 0.0001868
и 7.73е-05 0.0015523 0.0040031 0.00074	39 0.0036005 5.15e-05	0.0014525 0.0004831 0.0003382	0 0.0047277 0.0	0109675 0.0028051	0.0101543 0.0010467	0.0015587 0.0069853	0.0030885 0.0047921	1.93e-05 0.00038	0.0003349 0.0002	51 0.0026795 0.0019516	0.0006763 6.4e-06 3.22e-05	0 6.4e-06 3.86e-05
й 0.0012302 0 0	0 0.0039226	9 0.0010821	9 9	9 9	0 0.0048887	9 9	9 9	0.0002319	0	9 9 9	0 0.0023091 0	3.22e-05 3.86e-05 1.29e-05
K 0.0093556 0.0001256 0.0001997 7.41e-	05 0.0003671 0.0012174	4 0.0004026 7.73e-05 0.0029951	0.0001481 5.47e-05 0.6	0004734 0.0001965	0.0009533 0.0026408	0.0002383 0.000715	0.004615 0.0013269	0.0009758	0 0.0002	58 0.0004283 0.0017101	0 0.0002222 0.0021159	5.15e-05 1.29e-05 0.000541
л 0.0138965 0.0006795 0.0005056 0.00102	09 0.0010499 0.0089981	1 3.86e-05 0.0002641 0.0081672	0.0001159 0.0014847 0.6	0011787 7.73e-05	0 0.007755	0.0010724 0.0001159	0.0041996 0.0002963	0.0014364 2.58e-05	9.02e-05	0 6.4e-06 0.0009243	0 0.0045087 0	5.15e-05 2.58e-05 0.0015008
M 0.0048662 2.58e-05 0.0001159	0 8.05e-05 0.0051625	6.4e-06 0.0001288 0.0028791	0.0001224 0	3.2e-06 2.58e-05	0 0.0081576	0 0.000438	0.0013558 3.86e-05	0.0021803 6.4e-06	1.29e-05	0 3.2e-06 1.29e-05	0 0.0007472 0.0003446	1.29e-05 7.73e-05 0.0004895
н 0.003607 0.0002029 0.0012399 0.00021	58 0.0024701 0.0091495	0.0007697 0.0026666 0.0044636	0.0010531 0.0006538 0.	.000248 0.0037648	0.0020482 0.0114006	7.41e-05 0.0017487	0.0018389 0.001446	0.0002834 6	0.000219	0 0.0014492 0.0007182	1.93e-05 0.0002512 0.0013269	0.0002576 6.44e-05 0.0006763
0 6.44e-05 0.00257 0.0101028 0.01162	93 0.0050015 0.0001675	6.0001997 0.0004605 0.000409	0 0.0123314 0.0	0089434 0.0055039	0.0146888 0.0006248	0.0127339 0.0096905	0.0030981 0.0226113	1.61e-05 0.0002866	0.0041899 0.00026	73 0.0004187 0.0003897	6.4e-06 0 1.29e-05	0 0.0002061 0
n 0.0013204 0 0.0001192 1.93e-	05 3.54e-05 0.0008695	9.7e-06 0.0006312	2.58e-05 0 1.	.93e-05 0.000132	1.29e-05 0.0014041	3.86e-05 0.0001095	0.0026634 0.0001353	0.0006248	0	0 0 2.58e-05	0 0.0002287 1.29e-05	0 0 6.44e-05
p 0.0032173 0.0013462 0.0008889 0.00114	97 0.0019549 0.0093943	0 0.0002544 0.0007439	0 0.0024122	0 5.8e-05	0.0005668 0.0083927	0.0081769 7.41e-05	0.0005346 0.0042801	0.0011916 0.0001288	0.0001063	9 9 9		0.0001514 3.54e-05 6.4e-06
c 0.0056327 4.51e-05 0.0057325 1.93e-									0	9 9 9	0 0.0005604 0.0014041	1.29e-05 0.0005539 0.0002061
		1.29e-05 3.86e-05 0.0072204								0 0.0064797 0.0001868		0.0038711 0.0011819 0.0028115
y 0.0001481 0.0019838 0.0006924 0.00082			0 0.0033976 0.0	0018325 0.0028115	0.0026634 0.0001868	0.0008502 0.0035877	0.0006763 0.0025249	1.29e-05 0.0001031	0.0002287 0.0003	151 0.0013236 0.0003349	5.8e-05 0 0	0 0 0
φ 0.0001353 0 0 6.4e-	06 6.4e-06 0.0002319		0 00.0	0001031 0	1.29e-05 0.0002126	0 3.22e-05	1.93e-05 0	3.86e-05 1.29e-05	0	0 0 0	0 0.0001095	0 0 0
x 0.0009178 3.54e-05 1.93e-05	0 5.47e-05 0.0006409		0 6.4e-06		3.86e-05 0.0008309		0.0001578 1.29e-05		0	0 0 0	0 0.0009243 0	0 3.86e-05 0.0001095
ц 0.0002931 6.4e-06 1.93e-05	0 0.0003382 0.0005282				0.0006699 9.66e-05		3.54e-05 9.02e-05		0	0 0 0	0 6.4e-06 5.47e-05	0 0 3.22e-05
	06 1.93e-05 0.0015652				0.0002641 0.0033719		0.0004122 0.000161		0	0 6.4e-06 0	0 0.0001288 0.0002802	0 0.0001417 0.0002222
w 0.0018776 0 6.12e-05	0 0.0001031 0.0020354		3.22e-05 0 4.	.51e-05 6.4e-06	1.29e-05 0.0019774	1.29e-05 0.0003478	0.0001159 0	0.0014525 2.58e-05	6.4e-06	0 0.0002061 0	0 0.0006763 0.0006763	0 3.22e-05 6.4e-06
щ 0.0002126 0.0007761 1.93е-05	0 0.0010692		0 0		9.02e-05 0.0001578			9.98e-05 6	0	0 0 0	0 8.05e-05 6.4e-06	0 0.0001353 0.0003221
ы 0 0.0070046 0.0031046	0 0.0005571 0	0 0.0006795 0		.001417 0.0014589		0.000161 0.0014106		0 1.93e-05	0 0.0002		0 0 0	0 0 0
ь 0 0.0001578 0.00019		0 2.58e-05 0.0002061 0	0 00.0	0050916 0.000132	0.0023059 0	0.0002415 0.0004831	0.0037422 0.0112074	0 6	0	0 0.0004831 0.0018405	3.86e-05 0 0	0 0 0
9 0 0 1.29e-05	0 1.29e-05 0	0 0 2.58e-05	0 0	0 0	0 1.93e-05		0.0001224 6.4e-06		0	0 0 0	0 0 0	0 0 0
n 0.0026666 1.29e-05 1.29e-05	0 3.86e-05 0.0001578					6.4e-06 0.0011433			0	0 0 6.4e-06	0 0.0007633	0 2.58e-05 0.0002029
я 0.0021867 0.0010885 0.0002641	0 0.0002705 8.37e-05	0 0.0005346 0.0008856	0 00.0	0013816 0.0006505	0.0033204 0.0010434	0.0006699 0.0015909	0.0067405 0.0005475	1.29e-05 6	0	0 0 0	0 1.93e-05 0.0003929	0 0 1.61e-05

#### Таблиця частот символів:

### Frequency with spaces:

### Frequency with NO spaces

 $\label{eq:partial_constraints} $$\dot{\mathbf{n}}: 0.11543027977827465, \ '\mathbf{m}': 0.08654498156069845, \ '\mathbf{o}': 0.08373938448275428, \ '\mathbf{\phi}': 0.06896273572182997, \ '\mathbf{n}': 0.06631342527356462, \ '\mathbf{\kappa}': 0.06155676607672665, \ '\mathbf{s}': 0.05032429475658357, \ '\mathbf{r}': 0.050228506177103, \ '\mathbf{m}': 0.04125462873102819, \ '\mathbf{n}': 0.03947497775436279, \ '\mathbf{w}': 0.03641226396307602, \ '\mathbf{p}': 0.03060697189613493, \ '\mathbf{s}': 0.030422956993448564, \ '\mathbf{s}': 0.02968437663061151, \ '\mathbf{m}': 0.023884126067853603, \ '\mathbf{s}': 0.022737183866178314, \ '\mathbf{y}': 0.021774256567189388, \ '\mathbf{s}': 0.017907422858684116, \ '\mathbf{r}': 0.0177460947248221, \ '\mathbf{q}': 0.017551996813769358, \ '\mathbf{x}': 0.017226819793953723, \ '\mathbf{e}': 0.014373328426269262, \ '\mathbf{g}': 0.010866962266861941, \ '\mathbf{o}': 0.010093091374742569, \ '\mathbf{h}': 0.009422571318378552, \ '\mathbf{m}': 0.007544611010140986, \ '\mathbf{a}': 0.007272369784248828, \ '\mathbf{B}': 0.003977746800535408, \ '\mathbf{n}': 0.0028156800863105517, \ '\mathbf{n}': 0.002533355852052018, \ '\mathbf{c}': 0.001315832591812093$ 

#### Для символів:

H1 (spaces): 4.3283228

H1 (NO spaces): 4.4449543

Для Біграм:

H2 (intersection, spaces): 3.9242338

H2 (intersection, NO spaces): 4.0363426

H2 (NO intersection, spaces): 3.923236

H2 (NO intersection, NO spaces): 4.0363439

Розрахунок Н1 Н2 проводили за формулами

$$H(x_1, x_2, ..., x_n) = -\sum_{z_1, z_2, ..., z_n} P(x_1 = z_1, ..., x_n = z_n) \cdot \log_2 P(x_1 = z_1, ..., x_n = z_n).$$

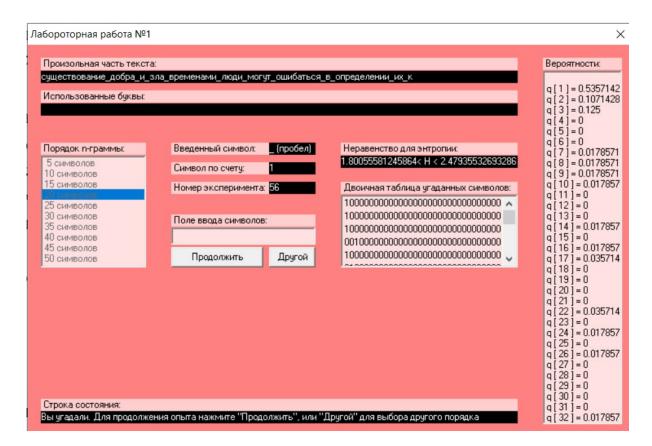
Ми проходили циклом по значенням частоти кожного символу(біграми) та домножували на логарифм основи двійки по ній - загальну суму з мінусом отримували як результат ентропії (для n-грами потрібно ще ділити на n)

## Оцінки для (10) Н, (20) Н, (30) Н (включно із відповідними скріншотами)

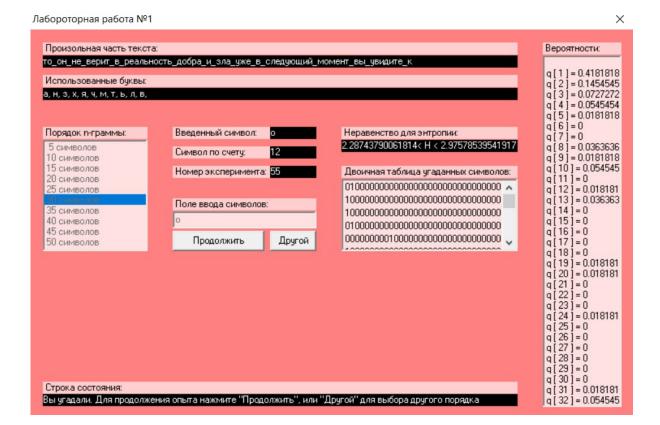
Оцінки для  $H^{(10)}$ ,  $H^{(20)}$ ,  $H^{(30)}$ умовних ентропій бралися з програми CoolPinkProgram шляхом 50+ експериментів з вгадування наступної літери в рядку

Лабороторная работа <b>№</b> 1			×
Произольная часть текста по_иными_ Использованные буквы: Ч, в, Порядок п-граммы: 5 символов 15 символов 20 символов 25 символов 30 символов 35 символов 40 символов 45 символов 50 символов	Введенный символ: в Символ по счету: 2 Номер эксперимента: 55 Поле ввода символов: Продолжить Другой	Неравенство для энтропии: 2.97414488795797< Н < 3.77954878919313  Двоичная таблица угаданных символов: 000000000000000000000100000000000 000000	Вероятности:  q[1] = 0.3333333 q[2] = 0.0925925 q[3] = 0.0370370 q[4] = 0.0185185 q[5] = 0.0740740 q[6] = 0.0185185 q[7] = 0.0185185 q[7] = 0.018518 q[17] = 0.018518 q[11] = 0 q[12] = 0.018518 q[13] = 0.018518 q[14] = 0.037037 q[15] = 0 q[16] = 0.018518 q[17] = 0 q[18] = 0.018518 q[17] = 0 q[18] = 0.018518 q[19] = 0.037037 q[20] = 0.018518 q[21] = 0 q[22] = 0.018518 q[24] = 0.018518 q[25] = 0.018518 q[27] = 0.018518 q[27] = 0.018518 q[27] = 0.018518 q[28] = 0 q[29] = 0.018518 q[29] = 0.018518
Строка состояния: Вы не угадали. Введите дру	угую букву		q[31]=0 q[32]=0

$$2.974 < H^{(10)} < 3.779$$



 $1.800 < H^{(20)} < 2.479$ 



 $2.287 < H^{(30)} < 2.975$ 

Верхні межі для умовної ентропії - 3.779, 2.479, 2.975

Нижні межі: 2.974, 1.800, 2.287

Отже верхня межа  $H1_{m} = 2.479$ , а нижня  $H2_{m} = 1.800$ 

## Оцінку надлишковості R російської мови у різних моделях відкритого тексту

За даними значеннями меж використовуючи формулу  $R=1-\frac{H_{\infty}}{H_0}$  ми розраховуємо надлишковість російської мови.  $H_{\infty}$  в даному випадку це ліміт куди прямує значення ентропії джерела символів. В наших експериментах  $H_{\infty}=3.923236$ , тому R=0.2153528

Межі зазначеної ентропії по экспериментам програми CoolPinkProgram при  $H1_{\infty}^{-}=$ ,  $H2_{\infty}^{-}=$  межі 0.50 < R < 0.63 (Значення такі низькі через відносно невелику кількість експериментів та людський фактор у виконанні завдання)

#### Висновки

В даній практичній роботі визначили ентропію російської мові на основі російського тексту через python. Обчислювалися частоти символів/біграм для розрахунку питомої ентропії на символ/біграму, визначення коефіцієнту хаотичності мови. Розібралися з переробкою письмових екземплярів в об'єкти аналізу частоти/ентропії. Ознайомилися з методикою програми CoolPinkProgram, що використовує принцип підрахунку умовної ентропії - як шанс вгадування людини наступних символів в контексті. Проаналізували наши труднощі й методи боротьби з ними. За нашим аналізом вийшло  $\pm$  0.215 і це значення відрізняється лише через людський фактор та не найбільший розмір текстового семплу мови.