

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

Виконав:  
студент групи ФБ-91  
Кузавка Артем

Київ – 2021

**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

Вихідний код створеної програми "frequencyCalculator.py" та зразки оброблюваних текстів розміщені на GitHub: [https://github.com/Yab0u/fb-labs-2021/tree/lab\\_1/cp\\_1](https://github.com/Yab0u/fb-labs-2021/tree/lab_1/cp_1)

Частоти символів у тексті з пробілами (зліва) та без них (справа):

	0.1681108754	о	0.1153326948
о	0.0959440145	е	0.0878266335
е	0.0730620213	а	0.0832081176
а	0.0692199281	н	0.0693391002
н	0.0576824434	и	0.0660166422
и	0.0549185267	т	0.0599351971
т	0.0498594387	с	0.0527814361
с	0.0439083027	л	0.0505470083
л	0.0420495065	в	0.0470651621
в	0.0391529965	р	0.0393950732
р	0.0327723329	к	0.0343133587
к	0.0285449099	д	0.0297711694
д	0.0247663121	м	0.028697357
м	0.0238730192	у	0.0272853123
у	0.0226983546	п	0.0241424466
п	0.0200838388	я	0.0217269299
я	0.0180743967	ь	0.0200320273
ь	0.0166644257	ы	0.0185429076
ы	0.0154256432	г	0.0181889611
г	0.0151311989	б	0.0176045377
б	0.0146450235	ч	0.0170799784
ч	0.0142086483	з	0.0163997725
з	0.0136427924	ж	0.0114901673
ж	0.0095585452	й	0.0102906402
й	0.0085606717	ш	0.0087042414
ш	0.0072409637	х	0.007660361
х	0.006372571	ю	0.0063006974
ю	0.0052414817	э	0.003565656
э	0.0029662304	щ	0.0028435452
щ	0.0023655143	ц	0.0025958574
ц	0.0021594655	ф	0.0010072136
ф	0.0008378901	ъ	0.000295579
ъ	0.000245889	ё	1.42E-05
ё	1.18E-05		

# Частоти біграм у тексті з пробілами та з кроком 1:

а	7E-06	0.001	0.003	0.001	0.002	0.001		0.001	0.004	0	0.001	0.004	0.01	0.002	0.004	4E-06	0.001	0.002	0.003	0.004	4E-05	0	0.001	0	0.001	0.001	0				1E-06	0.001	0.002
б	0.001	1E-06	9E-05	4E-06	3E-05	0.002		4E-06	2E-06	0.001		0	0.001	6E-05	0	0.002	1E-06	0.001	7E-05	8E-06	0.001		6E-05	6E-06	6E-06	8E-06	0	0	0.004	2E-05		4E-06	0.001
в	0.005	2E-05	2E-05	3E-05		0	0.004	1E-06	0.001	0.005		0	0.001	0	0.001	0.007	0	0.001	0.003	0	0.001		6E-05	2E-05	8E-05	0.001	8E-06	4E-06	0.002	0		2E-06	0
г	0.001		7E-06		0.001	0		1E-06	0.001			0	0.002	1E-06	0	0.008		0.001	3E-05	6E-06	0.001	1E-06			4E-05	6E-06				1E-06		2E-06	
д	0.004	2E-05	0.001	4E-06	4E-05	0.004		9E-06	1E-06	0.002		0	0.001	7E-05	0.002	0.004	7E-05	0.001	0	0	0.002		6E-05	0	3E-05	0	1E-06	6E-05	0.001	0.001		6E-06	0
е	3E-05	0.001	0.002	0.004	0.002	0.002		0.001	0.001	1E-04	0.002	0.002	0.006	0.004	0.007	0	0.001	0.005	0.004	0.004	9E-05	4E-06	0.001	0	0.001	0.001	0.001				1E-06	0	0
ё												1E-06																					
ж	0.001	4E-05		6E-06	0.001	0.004		1E-05		0.001		9E-05	4E-06	2E-06	0.001	4E-05		1E-06	1E-04		0			1E-06	4E-05					4E-05			
з	0.005	0	0.001	0.001	0.001	0		0	2E-05	0		9E-05	0	0	0.002	0		0	5E-05	1E-05	0			4E-06	6E-05	6E-06		1E-05	0	0		4E-06	0
и	8E-05	0	0.003	0	0.002	0.002		0	0.002	0.001	0.001	0.002	0.004	0.003	0.004	8E-05	0	0	0.002	0.004	9E-06	1E-05	0.001	0.001	0.002	0	0					0	0.002
й		1E-06			0	1E-06			1E-06			7E-05	9E-05	5E-05	0	2E-06		1E-06	0	0		4E-06	4E-05	8E-05	5E-05	1E-06							
к	0.008	1E-06	0		4E-06	0		8E-06	2E-06	0.003		1E-05	0		0.001	0.008		0.001	0.001	0	0.001	1E-06	1E-06	5E-06		1E-05							
л	0.008	2E-05	3E-06	0	1E-05	0.005		0	3E-06	0.005		0	0	1E-06	0	0.006	4E-05		0.001	2E-05	0.001	4E-06		1E-06	0	4E-06	1E-06		0.001	0.003		0.001	0.002
м	0.002	8E-06	1E-06	3E-05		0.003		1E-06	2E-06	0.003		5E-05	9E-05	4E-05	0.001	0.003	5E-05	7E-05	0	2E-05	0.003	1E-06	1E-06	9E-06	2E-05	4E-06	3E-06		0.001	6E-05		2E-06	0
н	0.011	1E-05	4E-06	7E-05	0.001	0.01		3E-06	1E-05	0.007		0		1E-06	0.003	0.01		8E-05	0.001	0	0.002	1E-05		0	0	2E-06	0		0.003	0.001		0	0.002
о	2E-06	0.003	0.007	0.005	0.004	0.002		0.002	0.001	0.001	0.003	0.002	0.005	0.005	0.008	0	0.001	0.005	0.006	0.006	5E-05	0	0	8E-05	0.002	0.001	0				3E-05	0.001	0
п	0.001					0.002				0.001		6E-05	0		6E-05	0.008	2E-05	0.006	3E-06	5E-05	0	1E-06		1E-05	3E-05	1E-05	1E-06		0	4E-05		3E-06	0
р	0.007	0	0	0	0	0.005		0	2E-05	0.004		0.001	5E-05	0	0.001	0.007	6E-05	2E-05	0	0	0.002	6E-06	1E-04	2E-05	7E-05	0	3E-05		0.001	0.001		0	0.001
с	0.002	5E-05	0.002	1E-05	0	0.004	1E-05	3E-05	2E-05	0.001		0.005	0.002	0.001	0.001	0.002	0.002	0	0.001	0.01	0	7E-06	0	5E-05	0	6E-05	2E-06	2E-05	0	0.003	2E-06	0	0.003
т	0.005	1E-05	0.003	6E-06	0	0.005		4E-06	2E-06	0.004		0	0	1E-05	0.001	0.014	5E-05	0.002	0.001	8E-05	0.001	2E-05	1E-05	6E-05	0	1E-06	1E-05	4E-05	0.001	0.006	1E-06	4E-05	0.001
у	3E-05	0.001	0.001	0.001	0.002	0		0.002	0	2E-06	0	0.001	0.001	0.001	0	4E-06	0.001	0.001	0.001	0.001	1E-06	2E-05	0	4E-06	0.001	0.001	0			1E-05	0.001	0	
ф	6E-05					0						0		2E-05	2E-06	1E-06	8E-05		0	9E-06	2E-05	2E-05	4E-06		1E-06				6E-06	5E-05		1E-06	1E-06
х	0.001		1E-04	1E-06	1E-06	3E-05				1E-04				9E-05	5E-05	1E-04	0.002	1E-06	6E-05	4E-05	6E-06	0			1E-06	9E-06		1E-06					
ц	0	1E-06	4E-05			0.001				0		0		1E-06			0					0		7E-06		1E-06			0				
ч	0.002		3E-06			0.003				0.001		0	4E-05	5E-06	0	5E-05		6E-05		0.004	0.001					0			0				
ш	0.001		9E-05	1E-06		0.002				0.002		0	0.001	6E-06	0	0	6E-06	1E-06		6E-05	0		2E-06		1E-06				0		2E-06		
щ	0					0.001				0.001						3E-05	1E-06		5E-06			9E-05							2E-05				
ъ						0																										7E-06	0
ы		0	0.001	9E-05	0	0.001		2E-05	6E-05	5E-06	0.001	0	0.003	0.001	0		0	0	0.001	0.001	6E-06		0.001	1E-06	0	0	1E-06					1E-06	
ь		8E-05	3E-05	5E-05	3E-05	0		1E-06	0	0		0.001		0	0.001	7E-06	2E-06		0.001	7E-05			1E-06	1E-04	6E-05	0	2E-05			2E-06	0	0	0
э			1E-06	3E-06	1E-06				3E-06	1E-06	2E-06	4E-05	4E-05	1E-06	2E-05		3E-06	4E-06	2E-06	0.003		2E-06	1E-06			1E-06							
ю	1E-06	0.001	3E-06	4E-06	0			2E-06	2E-06	1E-06		2E-06	1E-05	1E-05	3E-05		9E-06	4E-05	0	0		2E-06	1E-05	5E-06	4E-05	3E-05	0				3E-05		
я	1E-06	2E-05	0	0	0.001	7E-05		0	0	4E-05	0	0	0.001	0	0.001		6E-05	5E-05	0.001	0.001	1E-06		0	3E-05	0	5E-05	0				1E-04	0	

## Частоти біграм у тексті з пробілами та з кроком 2:

	а	б	в	г	д	е	ё	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я	
а	7E-06	0.001	0.003	0.001	0.002	0.001		0.001	0.004	0	0.001	0.004	0.01	0.002	0.004	4E-06	0.001	0.002	0.004	0.004	3E-05	0	0.001	0	0.001	0.001	0				1E-06	0.001	0.002	
б	0.001	2E-06	9E-05	4E-06	4E-05	0.002		4E-06	1E-06	0.001		0	0.001	6E-05	0	0.002	1E-06	0.001	8E-05	1E-05	0.001		6E-05	1E-05	4E-06	1E-05	0	0	0.004	1E-05	5E-06	0.001		
в	0.005	1E-05	2E-05	2E-05	0	0.004		1E-06	0.001	0.005		0	0.001	0	0.001	0.007	0	0.001	0.003	0	0.001		6E-05	2E-05	7E-05	0.001	5E-06	5E-06	0.002	0	2E-06	0		
г	0.001		6E-06		0.001	0		1E-06	0.001			0	0.002	1E-06	0	0.008		0.001	4E-05	7E-06	0.001				4E-05	9E-06					4E-06			
д	0.004	2E-05	0.001	4E-06	3E-05	0.005		1E-05		0.002		0	0.001	7E-05	0.001	0.004	6E-05	0.001	0	0	0.002		6E-05	0	3E-05	9E-05		7E-05	0.001	0.001	6E-06	0		
е	3E-05	0.001	0.002	0.004	0.002	0.002		0.001	0.001	9E-05	0.002	0.002	0.006	0.004	0.007	0	0.001	0.005	0.004	0.004	9E-05	4E-06	0.001	0	0.001	0.001	0.001				1E-06	0	0	
ё												1E-06																						
ж	0.001	5E-05		5E-06	0.001	0.004		1E-05		0.001		8E-05	4E-06	2E-06	0.001	5E-05			0		0				5E-05					4E-05				
з	0.005	0	0.001	0.001	0.001	0		0	2E-05	0		9E-05	0	0	0.002	0		0	5E-05	2E-05	0			7E-06	5E-05	9E-06		2E-05	0	0	7E-06	0		
и	8E-05	0	0.003	0	0.002	0.002		0	0.002	0.001	0.001	0.002	0.004	0.003	0.004	7E-05	0	0	0.002	0.004	1E-05	1E-05	0.001	0.001	0.002	0	0				0	0.002		
й		2E-06			0	1E-06						6E-05	9E-05	6E-05	0	2E-06			0	0		4E-06	4E-05	8E-05	5E-05	1E-06								
к	0.008	1E-06	0	4E-06	0			4E-06	2E-06	0.003		2E-05	0	0.001	0.008			0.001	0.001	0	0.001	1E-06		6E-06		1E-05								
л	0.008	2E-05	1E-06	0	6E-06	0.005		0	2E-06	0.005		0	0	1E-06	0	0.006	4E-05		0.001	2E-05	0.001	2E-06			0	1E-06			0.001	0.003	0.001	0.002		
м	0.002	1E-05		3E-05		0.003		1E-06	0.003			6E-05	9E-05	4E-05	0.001	0.003	5E-05	7E-05	0	2E-05	0.003	1E-06		6E-06	2E-05	4E-06	2E-06		0.001	7E-05	1E-06	0		
н	0.011	1E-05	5E-06	7E-05	0.001	0.01		2E-06	7E-06	0.007		0		0.003	0.01		8E-05	0.001	0	0.002	1E-05		0	0		0		0.003	0.001		0	0.002		
о	2E-06	0.003	0.007	0.005	0.004	0.002		0.002	0.001	0.001	0.003	0.001	0.005	0.005	0.008	0	0.001	0.005	0.006	0.006	4E-05	0	0	8E-05	0.002	0.001	0			3E-05	0.001	0		
п	0.001					0.002				0.001		7E-05	0	6E-05	0.008	2E-05	0.006	4E-06	4E-05	0	1E-06		1E-05	3E-05	2E-05			0	4E-05	2E-06	0			
р	0.007	0	0	0	0	0.005		0	2E-05	0.004		0.001	6E-05	0	0.001	0.007	6E-05	1E-05	0	0	0.002	4E-06	9E-05	2E-05	6E-05	0	3E-05		0.001	0.001		0	0.001	
с	0.002	5E-05	0.002	1E-05	0	0.004	1E-05	2E-05	2E-05	0.001		0.005	0.002	0.001	0.001	0.002	0.002	0	0.001	0.01	0	7E-06	0	5E-05	0	5E-05	4E-06	2E-05	0	0.003	1E-06	0	0.003	
т	0.005	1E-05	0.003	7E-06	0	0.005		4E-06	2E-06	0.004		0	0	2E-05	0.001	0.014	5E-05	0.002	0.001	8E-05	0.001	2E-05	1E-05	6E-05	0		1E-05	4E-05	0.001	0.006	1E-06	5E-05	0	
у	3E-05	0	0.001	0.001	0.002	0		0.002	0	2E-06	0	0.001	0.001	0.001	0	4E-06	0.001	0.001	0.001	0.001		1E-05	0	4E-06	0.001	0.001	0			1E-05	0.001	0		
ф	6E-05				1E-04					0		2E-05	1E-06		9E-05		0	1E-05	2E-05	2E-05	9E-06			1E-06				5E-06	5E-05		2E-06	1E-06		
х	0.001		1E-04	1E-06	4E-05			9E-05				9E-05	6E-05	1E-04	0.002		6E-05	4E-05	7E-06	0				1E-06	1E-05		1E-06							
ц	0	1E-06	4E-05		0.001					0		9E-05	1E-06			0					0		6E-06						0					
ч	0.002		2E-06		0.003					0.001		0	4E-05	5E-06	0	5E-05		6E-05		0.005	0.001				0				0					
ш	0.001		9E-05	1E-06		0.002				0.002		0	0.001	5E-06	0	0	7E-06			7E-05	0			4E-06		1E-06				0.001	2E-06			
щ	0					0.001				0.001					3E-05	1E-06		5E-06			8E-05									2E-05				
ъ						0																										7E-06	0	
ы		0	0.001	9E-05	0	0.001		3E-05	6E-05	4E-06	0.001	0	0.003	0.001	0		0	0	0.001	0.001	7E-06		0.001		0	0						1E-06		
ь		8E-05	3E-05	4E-05	2E-05	0		1E-06	0	0		0.001		0	0.001	7E-06	4E-06		0.001	8E-05			1E-06	9E-05	6E-05	0	2E-05			2E-06	0	0		
э			1E-06	4E-06				4E-06		4E-06	4E-05	4E-05	2E-06	3E-05		4E-06	5E-06	4E-06	0.003			1E-06												
ю	1E-06	0.001	4E-06	4E-06	0			2E-06	1E-06	1E-06		1E-06	1E-05	1E-05	3E-05		1E-05	3E-05	0	0		2E-06	1E-05	5E-06	4E-05	3E-05	0				3E-05			
я		2E-05	0	0	0.001	6E-05		0	0	4E-05	0	0	0.001	0	0.001		6E-05	5E-05	0.001	0.001			0	3E-05	0	5E-05	0					0	0	

# Частоти біграм у тексті без пробілів та з кроком 1:

	а	б	в	г	д	е	ё	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я	
а	0.001	0.002	0.006	0.001	0.004	0.002		0.002	0.006	0.002	0.001	0.006	0.012	0.004	0.008	0.002	0.002	0.003	0.006	0.006	0.001	0	0.001	0	0.002	0.001	0					0	0.001	0.003
б	0.001	4E-06	0	8E-06	4E-05	0.002		4E-06	4E-06	0.001		0	0.001	7E-05	0	0.002	7E-06	0.001	9E-05	1E-05	0.001		7E-05	8E-06	8E-06	1E-05	0	0	0.005	2E-05	0	5E-06	0.001	
в	0.007	0	0	0	0.001	0.005		6E-05	0.001	0.006		0.001	0.001	0	0.002	0.009	0.001	0.001	0.005	0.001	0.001	1E-05	1E-04	5E-05	0	0.001	1E-05	5E-06	0.003	0	0	4E-06	0	
г	0.001	4E-05	9E-05	9E-06	0.001	0		2E-05	3E-05	0.001		0	0.002	2E-05	0	0.01	0	0.001	0	3E-05	0.001	3E-06	2E-06		5E-05	8E-06				1E-06	1E-05	2E-06	6E-06	
д	0.005	8E-05	0.001	3E-05	8E-05	0.005		3E-05	4E-05	0.003		0	0.001	0	0.002	0.004	0	0.002	0.001	0	0.002	2E-06	8E-05	0	8E-05	0	1E-06	7E-05	0.001	0.001	3E-05	7E-06	0	
е	0	0.003	0.005	0.005	0.004	0.003		0.001	0.002	0.002	0.003	0.003	0.007	0.006	0.01	0.002	0.004	0.007	0.007	0.006	0.001	5E-05	0.001	0	0.002	0.001	0.001			1E-06	0	0	0.001	
ё			2E-06			1E-06					1E-06		1E-06		1E-06	1E-06	1E-06		1E-06		1E-06		1E-06								1E-06			
ж	0.002	7E-05	2E-05	8E-06	0.001	0.005		2E-05	8E-06	0.002		0	9E-06	1E-05	0.001	7E-05	3E-05	5E-06	0	3E-05	0			1E-06	5E-05	1E-06					5E-05	1E-05		2E-05
з	0.007	0	0.001	0.001	0.001	0		0	9E-05	0		0	0	0	0	0.002	0	0	0	0	9E-05	0	2E-06	4E-06	9E-06	9E-05	2E-05		2E-05	0	0	4E-05	5E-06	0.001
и	0	0.001	0.005	0.001	0.003	0.003	1E-06	0.001	0.003	0.002	0.002	0.003	0.006	0.004	0.007	0.002	0.002	0.001	0.005	0.006	0.001	7E-05	0.002	0.001	0.003	0.001	0				0	0	0.002	
й	0	0	0.001	0	0.001	0		0	0	0.001		0.001	0	0	0.001	0.001	0.001	0	0.001	0.001	0	3E-05	8E-05	7E-05	0	0	1E-05				7E-05	2E-06	7E-05	
к	0.009	0	0.001	1E-04	0	0.001		0	0	0.004		0	0	0	0.001	0.01	0	0.002	0.002	0.001	0.002	1E-05	9E-05	1E-05	0	3E-05	9E-06				0	1E-06	0	
л	0.009	0	0.001	0	0	0.007		0	0	0.007		0.001	0.001	0	0.001	0.008	0.001	0	0.002	0	0.002	2E-05	4E-05	2E-05	0.001	5E-05	4E-06		0.001	0.004	0	0.001	0.002	
м	0.003	0	0.001	0	0	0.004		0	0	0.004		0.001	0	0	0.002	0.004	0.001	0	0.001	0	0.004	2E-05	9E-05	2E-05	0.001	7E-05	8E-06		0.001	8E-05	8E-05	5E-06	0.001	
н	0.013	0	0.001	0	0.001	0.012		7E-05	0	0.009		0.001	1E-04	0	0.004	0.012	0.001	0	0.002	0.001	0.003	2E-05	1E-04	0	0.001	2E-05	0		0.003	0.001	5E-05	0	0.002	
о	0	0.006	0.011	0.006	0.006	0.004		0.003	0.002	0.002	0.003	0.003	0.007	0.007	0.013	0.003	0.004	0.007	0.01	0.009	0.001	0	0.001	0	0.003	0.002	0				0.001	0.001	0.001	
п	0.001	1E-06	4E-06		1E-06	0.002		1E-06	1E-06	0.001		7E-05	0.001	1E-06	8E-05	0.01	3E-05	0.007	1E-05	6E-05	0.001	1E-06		1E-05	4E-05	2E-05	1E-06		0	4E-05	1E-06	4E-06	0	
р	0.008	0	0	0	0	0.006		0	4E-05	0.005		0.001	7E-05	0	0.001	0.008	0	3E-05	0	0.001	0.003	1E-05	0	3E-05	9E-05	0	4E-05		0.001	0.001	4E-06	0	0.001	
с	0.002	0	0.002	0	0.001	0.005	1E-05	0	9E-05	0.002		0.006	0.003	0.001	0.002	0.003	0.002	0	0.001	0.012	0.001	2E-05	0	6E-05	0.001	8E-05	4E-06	3E-05	0	0.003	6E-05	0	0.004	
т	0.006	0	0.004	8E-05	0	0.006		0	0	0.005		0.001	0	0	0.002	0.017	0	0.003	0.001	0	0.001	4E-05	5E-05	8E-05	0.001	2E-05	2E-05	4E-05	0.002	0.008	0	5E-05	0.001	
у	0	0.001	0.002	0.001	0.002	0		0.002	0	0.001	0	0.001	0.002	0.002	0.001	0.001	0.001	0.001	0.002	0.002	0	5E-05	0	1E-05	0.002	0.001	0				9E-05	0.001	0	
ф	8E-05	3E-06	7E-06	1E-06		0					0	1E-06	2E-05	3E-06	4E-06	9E-05	2E-06	0	2E-05	2E-05	3E-05	5E-06			2E-06				7E-06	6E-05		1E-06	1E-06	
х	0.001	0	0	8E-05	0	0		5E-05	6E-05	0		0	0	0	0	0.003	0	0	0	0	0	9E-06	2E-05	1E-05	6E-05	3E-05	2E-06	1E-06			3E-05	2E-06	3E-05	
ц	0.001	1E-05	7E-05	6E-06	9E-06	0.001		2E-06	4E-06	0		0	1E-06	7E-06	2E-05	0	2E-05	4E-06	2E-05	8E-06	0		1E-06	1E-05	4E-06	1E-06		0		1E-06		4E-06		
ч	0.003	3E-05	0	1E-05	4E-05	0.004		6E-06	4E-05	0.002		0	6E-05	2E-05	0.001	0	0	9E-05	1E-04	0.005	0.001		9E-06	1E-06	2E-05	0			0	4E-06		2E-05		
ш	0.001	7E-06	0	3E-06	1E-06	0.003		1E-06	1E-06	0.002		0	0.001	1E-05	0	0	2E-05	4E-06	4E-06	8E-05	0		4E-06	3E-06	4E-06	3E-06				0.001	1E-06	3E-06	1E-06	
щ	0		1E-06			0.002				0.001			1E-06	1E-06	3E-05	2E-06	2E-06	6E-06	1E-06		0									3E-05	1E-06			
ъ						0		1E-06								1E-06																8E-06	0	
ы	8E-05	0.001	0.002	0	0	0.001		9E-05	0	0	0.001	0	0.003	0.002	0.001	0	0.001	0	0.001	0.001	0	1E-05	0.001	7E-06	0	0.001	4E-06				5E-05		6E-05	
ь	0	0	0.001	0	0.001	0.001		0	0.001	0.001		0.002	0	0.001	0.003	0.001	0.001	0	0.002	0.001	0	2E-05	0	0	0.001	0.001	2E-05				0	0	0.001	
э			1E-06	4E-06	1E-06				4E-06	1E-06	3E-06	5E-05	4E-05	1E-06	3E-05		4E-06	5E-06	3E-06	0.003		2E-06	1E-06			1E-06								
ю	0	0.001	0	0	0	0		0	1E-04	0		0	0	0	0	0	0	0	0	0.001	0	0	3E-05	4E-05	2E-05	0	7E-05	0			7E-05	4E-05	6E-05	
я	0	0.001	0.001	0.001	0.001	0.001		0.001	0.001	0.001	0	0.001	0.001	0.001	0.003	0.001	0.001	0	0.002	0.002	0	3E-05	0	6E-05	0.001	0	0				0	0		

## Частоти біграм у тексті без пробілів та з кроком 2:

	а	б	в	г	д	е	ё	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я	
а	0.001	0.002	0.006	0.001	0.004	0.002		0.002	0.005	0.002	0.001	0.006	0.012	0.004	0.008	0.002	0.002	0.003	0.006	0.006	0.001	0	0.001	0	0.002	0.001	0					0	0.001	0.003
б	0.001	4E-06	0	9E-06	5E-05	0.002		4E-06	6E-06	0.001		0	0.001	6E-05	0	0.003	6E-06	0.001	9E-05	2E-05	0.001		7E-05	7E-06	7E-06	1E-05	0	0	0.005	2E-05	1E-04	6E-06	0.001	
в	0.007	0	0	0	0.001	0.005		6E-05	0.001	0.006		0.001	0.001	0	0.002	0.009	0.001	0.001	0.005	0.001	0.001	2E-05	0	5E-05	0	0.001	1E-05	4E-06	0.003	0	0	7E-06	0	
г	0.001	5E-05	8E-05	7E-06	0.001	0		2E-05	2E-05	0.001		0	0.002	1E-05	0	0.01	1E-04	0.001	0	3E-05	0.001	1E-06	3E-06		6E-05	1E-05				1E-06	1E-05		1E-05	
д	0.005	7E-05	0.001	3E-05	6E-05	0.005		2E-05	5E-05	0.002		0	0.001	0	0.002	0.004	0	0.002	0.001	0	0.002	3E-06	8E-05	0	8E-05	0	1E-06	8E-05	0.001	0.001	3E-05	1E-05	0	
е	0	0.003	0.005	0.005	0.004	0.003		0.001	0.002	0.002	0.003	0.003	0.008	0.006	0.01	0.002	0.004	0.007	0.007	0.006	0.001	6E-05	0.001	0	0.002	0.001	0.001				0	0.001	0.001	
ё			3E-06							3E-06				1E-06	1E-06	1E-06	1E-06						1E-06								1E-06			
ж	0.002	7E-05	2E-05	7E-06	0.001	0.005		2E-05	9E-06	0.002		0	7E-06	7E-06	0.001	8E-05	3E-05	6E-06	0	2E-05	0			1E-06	5E-05	1E-06				5E-05	1E-05		2E-05	
з	0.007	0	0.001	0.001	0.001	0		0	8E-05	0		0	0	0	0	0.002	0	0	0	0	9E-05	0	1E-06	1E-06	1E-05	0	2E-05		2E-05	0	0	5E-05	7E-06	0.001
и	0	0.001	0.005	0.001	0.003	0.003	1E-06	0.001	0.003	0.002	0.002	0.003	0.006	0.004	0.007	0.002	0.002	0.001	0.005	0.006	0.001	6E-05	0.002	0.001	0.003	0.001	0				0	0	0.002	
й	0	0	0.001	0	0.001	0		0	0	0.001		0.001	0	0	0.001	0.001	0.001	0	0.001	0.001	0	2E-05	7E-05	7E-05	0	0	1E-05				7E-05	3E-06	6E-05	
к	0.009	0	0.001	0	0	0.001		0	0	0.004		0	0	0	0.002	0.01	0	0.002	0.002	0.001	0.002	1E-05	0	2E-05	0	4E-05	7E-06				0	1E-06	0	
л	0.009	0	0.001	0	0	0.007		0.001	0	0.007		0.001	0.001	0	0.001	0.008	0.001	0	0.002	0	0.002	2E-05	5E-05	2E-05	0.001	3E-05	4E-06		0.001	0.004	0	0.001	0.002	
м	0.003	0	0.001	0	0	0.004		0	0	0.004		0.001	0	0	0.002	0.004	0.001	0	0.001	0	0.003	2E-05	9E-05	2E-05	0.001	8E-05	9E-06		0.001	8E-05	8E-05	6E-06	0.001	
н	0.013	0	0.001	0	0.001	0.012		6E-05	0	0.009		0.001	0	0	0.004	0.012	0.001	0	0.002	0.001	0.003	2E-05	7E-05	0	0.001	3E-05	0		0.003	0.001	5E-05	0	0.002	
о	0	0.005	0.011	0.006	0.006	0.004		0.003	0.002	0.002	0.004	0.003	0.007	0.007	0.013	0.003	0.004	0.007	0.009	0.01	0.001	0	0.001	0	0.003	0.002	0				0.001	0.001	0.001	
п	0.001	1E-06			1E-06	0.002		1E-06	3E-06	0.001		7E-05	0.001	3E-06	8E-05	0.01	2E-05	0.007	2E-05	6E-05	0.001	1E-06		1E-05	4E-05	2E-05				0	6E-05	1E-06	0	
р	0.008	0	0	0	0	0.006		0	3E-05	0.005		0.001	6E-05	0	0.001	0.008	0	4E-05	0	0.001	0.003	2E-05	0	3E-05	9E-05	0	3E-05		0.001	0.001	4E-06	0	0.001	
с	0.002	0	0.002	0	0.001	0.005	1E-05	0	8E-05	0.002		0.006	0.003	0.001	0.002	0.003	0.002	0	0.001	0.012	0.001	2E-05	0	6E-05	0.001	7E-05	6E-06	2E-05	0	0.003	6E-05	0	0.004	
т	0.006	0	0.004	8E-05	0	0.006		1E-04	0	0.005		0.001	0	0	0.002	0.018	0	0.003	0.001	0	0.002	4E-05	6E-05	8E-05	0.001	9E-06	2E-05	5E-05	0.002	0.008	0	5E-05	0.001	
у	0	0.001	0.002	0.001	0.002	0		0.002	0.001	0.001	0	0.001	0.002	0.002	0.001	0.001	0.001	0.001	0.002	0.002	0	5E-05	0	1E-05	0.002	0.001	0				9E-05	0.001	0	
ф	9E-05	4E-06	9E-06	1E-06		0					0	1E-06	2E-05		4E-06	9E-05		0	2E-05	3E-05	3E-05	7E-06			3E-06				3E-06	6E-05	3E-06	1E-06		
х	0.001	0	0	8E-05	0	0		5E-05	7E-05	0		0	0	0	0	0.003	0	0	0	1E-04	0	1E-05	2E-05	1E-05	8E-05	2E-05	3E-06	1E-06			3E-05	1E-06	4E-05	
ц	0.001	6E-06	5E-05	6E-06	1E-05	0.001		1E-06	6E-06	0		0	1E-06	4E-06	2E-05	0	2E-05	4E-06	2E-05	7E-06	0		1E-06	1E-05		1E-06		0				6E-06		
ч	0.003	3E-05	0	1E-05	4E-05	0.004		6E-06	3E-05	0.002		0	6E-05	2E-05	0.001	0	0	9E-05	0	0.005	0.001		1E-05	1E-06	2E-05	0			0	9E-06		2E-05		
ш	0.001	4E-06	9E-05	3E-06	1E-06	0.003		1E-06		0.002		0	0.001	2E-05	0	0	1E-05	3E-06	3E-06	8E-05	0		3E-06	3E-06	3E-06	3E-06				0.001		4E-06		
щ	0		1E-06			0.002				0.001			1E-06	4E-05	1E-06	4E-06	6E-06				0									3E-05				
ъ						0																										1E-05	0	
ы	8E-05	0.001	0.002	0	0	0.001		9E-05	0	0	0.001	0	0.003	0.002	0.001	0	0.001	0	0.001	0.001	9E-05	1E-05	0.001	6E-06	0	0.001	7E-06				5E-05		6E-05	
ь	0	0	0.001	0	0	0.001		0	0.001	0.001		0.002	0	0.001	0.003	0.001	0.001	0	0.002	0.001	0	3E-05	9E-05	0	0.001	0.001	3E-05				0	0	0.001	
э				3E-06	3E-06				3E-06	1E-06	3E-06	4E-05	4E-05		3E-05		3E-06	3E-06	3E-06	0.003						1E-06								
ю	0	0.001	0	1E-04	0	0		0	0	0		0	9E-05	0	0	0	0	0	0	0.001	0	0	2E-05	5E-05	3E-05	0	7E-05	0			8E-05	4E-05	5E-05	
я	0	0.001	0.001	0.001	0.001	0.001		0.001	0.001	0.001	0	0.001	0.001	0.001	0.003	0.001	0.001	0	0.002	0.002	0	3E-05	0	7E-05	0.001	0	0				0	0	0	

За отриманими даними була розрахована питома ентропія на символ та на символ біграми:

Текст із пробілами:

$$H_1 = 4.3471$$

$$H_2 = 3.9375$$

Для біграм з кроком 2:

$$H_2 = 3.9373$$

Текст без пробілів:

$$H_1 = 4.4401$$

$$H_2 = 4.1197$$

Для біграм з кроком 2:

$$H_2 = 4.1194$$



Розрахунок умовної ентропії джерела:

$H^{(50)}$

Лабораторная работа №1

Произвольная часть текста:  
\_человеческие\_существа\_во\_всех\_частях\_земного\_шар

Использованные буквы:

Порядок n-граммы:  
5 символов  
10 символов  
15 символов  
20 символов  
25 символов  
30 символов  
35 символов  
40 символов  
45 символов  
50 символов

Введенный символ:  
Символ по счету:  
Номер эксперимента: 50

Поле ввода символов:  

Продолжить Другой

Неравенство для энтропии:  
1,67350951985244< H < 2,25964913004547

Двоичная таблица угаданных символов:  
10000000000000000000000000000000  
10000000000000000000000000000000  
10000000000000000000000000000000  
01000000000000000000000000000000  
00000000000000000000000000000000

Вероятности:  
q[ 1 ] = 0,5510204  
q[ 2 ] = 0,1428571  
q[ 3 ] = 0,1020408  
q[ 4 ] = 0,0204081  
q[ 5 ] = 0  
q[ 6 ] = 0  
q[ 7 ] = 0  
q[ 8 ] = 0  
q[ 9 ] = 0  
q[ 10 ] = 0  
q[ 11 ] = 0,020408  
q[ 12 ] = 0  
q[ 13 ] = 0  
q[ 14 ] = 0,020408  
q[ 15 ] = 0  
q[ 16 ] = 0  
q[ 17 ] = 0,020408  
q[ 18 ] = 0,020408  
q[ 19 ] = 0  
q[ 20 ] = 0  
q[ 21 ] = 0  
q[ 22 ] = 0,061224  
q[ 23 ] = 0  
q[ 24 ] = 0  
q[ 25 ] = 0  
q[ 26 ] = 0,020408  
q[ 27 ] = 0,020408  
q[ 28 ] = 0  
q[ 29 ] = 0  
q[ 30 ] = 0  
q[ 31 ] = 0  
q[ 32 ] = 0

Строка состояния:

$H^{(30)}$

Лабораторная работа №1

Произвольная часть текста:  
ия\_как\_все\_организмы\_подчиняю

Использованные буквы:

Порядок n-граммы:  
5 символов  
10 символов  
15 символов  
20 символов  
25 символов  
30 символов  
35 символов  
40 символов  
45 символов  
50 символов

Введенный символ:  
Символ по счету:  
Номер эксперимента: 50

Поле ввода символов:  

Продолжить Другой

Неравенство для энтропии:  
2,17775849810142< H < 2,66131527740124

Двоичная таблица угаданных символов:  
00010000000000000000000000000000  
00010000000000000000000000000000  
00000000000000000000000000000001  
10000000000000000000000000000000  
00000000000000000000000000000010

Вероятности:  
q[ 1 ] = 0,44  
q[ 2 ] = 0,12  
q[ 3 ] = 0,08  
q[ 4 ] = 0,12  
q[ 5 ] = 0  
q[ 6 ] = 0  
q[ 7 ] = 0,04  
q[ 8 ] = 0  
q[ 9 ] = 0,04  
q[ 10 ] = 0  
q[ 11 ] = 0  
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 ] = 0,04  
q[ 22 ] = 0  
q[ 23 ] = 0,04  
q[ 24 ] = 0  
q[ 25 ] = 0  
q[ 26 ] = 0  
q[ 27 ] = 0  
q[ 28 ] = 0  
q[ 29 ] = 0  
q[ 30 ] = 0  
q[ 31 ] = 0,04  
q[ 32 ] = 0,04

Строка состояния:

$H^{(20)}$ 

Лабораторная работа №1

Произвольная часть текста:  
тенья\_или\_наследств

Использованные буквы:

Порядок n-граммы:  
5 символов  
10 символов  
15 символов  
20 символов  
25 символов  
30 символов  
35 символов  
40 символов  
45 символов  
50 символов

Введенный символ:

Символ по счету:

Номер эксперимента: 50

Поле ввода символов:

Продолжить

Другой

Неравенство для энтропии:  
3,80597090931256 < H < 3,5804621997271

Двоичная таблица угаданных символов:

00000000000000000000000000000000  
00000000000000000000000000000000  
00000000000000000000000000000000  
00010000000000000000000000000000  
00000000000000000000000000000000

Вероятности:

q[1] = 0,1923076  
q[2] = 0,0769230  
q[3] = 0,0384615  
q[4] = 0,1153846  
q[5] = 0  
q[6] = 0,0384615  
q[7] = 0  
q[8] = 0,0384615  
q[9] = 0,0384615  
q[10] = 0  
q[11] = 0  
q[12] = 0  
q[13] = 0  
q[14] = 0  
q[15] = 0  
q[16] = 0,0769230  
q[17] = 0  
q[18] = 0  
q[19] = 0,0769230  
q[20] = 0,1153846  
q[21] = 0  
q[22] = 0  
q[23] = 0,0384615  
q[24] = 0,0769230  
q[25] = 0,0384615  
q[26] = 0  
q[27] = 0  
q[28] = 0  
q[29] = 0  
q[30] = 0  
q[31] = 0  
q[32] = 0,0384615

Строка состояния:

 $H^{(10)}$ [illegible]

Оцінка надлишковості російської мови:

- Для тексту з пробілами та з буквами "ъ" та "ё"  
 $R = 0,46$
- Для тексту з пробілами та без букв "ъ" та "ё"  
 $R = 0,455$
- Для тексту без пробілів та з буквами "ъ" та "ё"  
 $R = 0,38$
- Для тексту без пробілів та без букв "ъ" та "ё"  
 $R = 0,37$

Контрольні запитання:

1. Дайте визначення ентропії, сукупної ентропії, умовної ентропії. Який зміст несуть ці поняття?  
Ентропія – міра місткості об'єкта щодо інформації.  
Сукупна ентропія – міра місткості двох ансамблів щодо інформації.  
Умовна ентропія – міра місткості одного ансамблю щодо інформації, якщо поведінка іншого є визначеною.
2. Які визначення ентропії на символ джерела ви знаєте?  
$$H_{\infty} = \lim_{n \rightarrow \infty} H_n, \text{ де } H_n = \frac{1}{n} H(x_1, x_2, \dots, x_n),$$
3. Порівняйте одержані значення  $H_1$ ,  $H_2$ ,  $H^{(10)}$ ,  $H^{(20)}$ ,  $H^{(30)}$ . Зробіть висновки.  
Щодо  $H_1$ ,  $H_2$  можна зробити висновок, що питома ентропія на символ та на символ біграми менша, ніж на окремий символ. Щодо умовної ентропії, наглядно, що ми маємо менше інформації зі зменшенням кількості попередніх символів перед тим, який потрібно вгадати, тому це зробити складніше.
4. Що таке надлишковість джерела? Яка надлишковість російського письмового тексту згідно ваших даних?  
Надлишковість джерела – кількість інформації, яку можна усунути без втрати сенсу інформації в джерелі. Обчислена надлишковість російської мови приблизно дорівнює 4,2.
5. Які моделі відкритих текстів розглядаються у криптографії?  
Моделі, що відрізняються залежно від наявності пробілів, розділових знаків, деяких букв алфавіту, якими можна знехтувати та в залежності від регістру літер.