Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» Фізико-технічний інститут

КРИПТОГРАФІЯ КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

> Виконали: ФБ-31 Федорович Дарина ФБ-31 Шваюк Олександра

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:

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.

Хід роботи:

Для аналізу використали текст: https://gist.github.com/romaklimenko/c95f3a864828f7f034b7a33d1676e420

Спочатку проводимо підготовку тексту до аналізу, переведення усього тексту у нижній регістр, а також видалення всіх лишніх символів, крім літер російського алфавіту та пробілів. Цей аналіз проводимо двічі один раз з урахуванням пробілів, а інший без них для порівняння результату. Пробіл замінили на «Д.».

Наступним кроком є підрахунок частот п-грам. Програма підраховує кількість кожнії монограми та кожної біграми. Підрахунок біграм здійснюється з перетином та без. Обраховуємо ентропію монограм H1 та біграм H2 та обраховуємо надлишковість R.

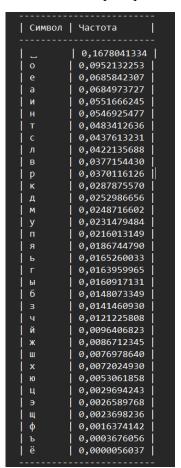
Усі отримані результати записуємо у файл "crypto analysis.txt"

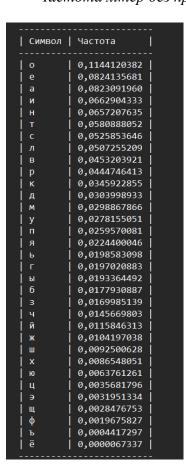
Запускаємо код:

```
Аналіз завершено. Результати збережено у 'crypto_analysis.txt'
Process finished with exit code 0
```

Частота літер з пробілами:







Біграми, що перетиняються без пробілів:

0,0057 0,0027 0,0075 0,0148 0,0047 0,0072 0,0042 0,0020 0,0039 0,0113 0,0000 0,0102 0,0026 0,0056 0,0154 0,0120 0,0152 0,0039 0,0155 0,0068 0,0039 0,0015 0,0003 0,0055 0,0008 0,0001 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0015 0,0000 0,0162 0,0000 0,0006 0,0031 0,0008 0,0020 0,0008 0,0011 0,0036 0,0001 0,0008 0,0044 0,0088 0,0027 0,0047 0,0000 0,0008 0,0028 0,0035 0,0048 0,0001 0,0005 0,0010 0,0007 0,0012 0,0002 0,0000 0,0000 0,0000 0,0000 0,0008 0,0024 0,0000 0,0002 0,0010 0,0000 0,0001 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0001 0,0000 0,0 0,0060 0,0054 0,0000 0,0000 0,0000 0,0000 0,0002 0,0004 0,0000 0,0005 0,0000 0,0 0,0007 0,0000 0,0 0,0010 0,0041 0,0000 0,0000 0,0000 0,0000 0,0000 0,0004 0,0000 0,0004 0,0000 0,0004 0,0000 0,0001 0,0001 0,0001 0,0001 0,0001 0,0001 0,0000 0,0001 0,0 0,0169 0,0000 0,0011 0,0013 0,0034 0,0024 0,0016 0,0010 0,0012 0,0001 0,0024 0,0010 0,0010 0,0001 0,0000
0,0000 0,0 0,0002 0,0013 0,0001 0,0000 0,0 0,0011 0,0050 0,0001 0,0007 0,0004 0,0007 0,0003 0,0001 0,0001 0,0003 0,0001 0,0000 0,0 0,0187 0,0001 0,0004 0,0024 0,0005 0,0016 0,0023 0,0003 0,0017 0,0007 0,0013 0,0018 0,0018 0,0028 0,0030 0,0002 0,0002 0,0002 0,0002 0,0000 0,0000 0,0000 0,0000 0,0001 0,0001 0,0000 0,0 0,0078 0,0000 0,0 0,0048 0,0069 0,0000
0,0000 0,0 0,0079 0,0067 0,0000 0,0001 0,0002 0,00037 0,0003 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0, 0,0074 0,0027 0,0000 0,0 0,0046 0,0101 0,0000 0,0001 0,0001 0,0001 0,0001 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0001 0,0000 0,0000 0,0001 0,0000 0,0000 0,0000 0,0001 0,0000 0,0 0,0216 0,0000 0,0031 0,0074 0,0045 0,0043 0,0022 0,0016 0,0008 0,0008 0,0008 0,0005 0,0050 0,0068 0,0002 0,0010 0,0062 0,0067 0,0001 0,0003 0,0006 0,0001 0,0013 0,0012 0,0001 0,0000 0,0000 0,0000 0,0000 0,0005 0,0006 0,000 0,0000 0,0012 0,0000 0,0 0,0013 0,0078 0,0001 0,0003 0,0004 0,0004 0,0001 0,0003 0,0000 0,0004 0,0001 0,0000 0,0002 0,0001 0,0002 0,0000 0,0001
0,0001 0,0 0,0029 0,0014 0,0000 0,0018 0,0000 0,0003 0,0006 0,0000 0,0000 0,0004 0,0000 0,0004 0,0000 0,0004 0,0001 0,0000 0,0002 0,0000 0,0 0,0044 0,0055 0,0000 0,0024 0,0000 0,0002 0,0004 0,0000 0,0002 0,0004 0,0000 0,0 0,0062 0,0001 0,0005 0,0008 0,0012 0,0018 0,0002 0,0013 0,0007 0,0000 0,0001 0,0008 0,0013 0,0001 0,0000 0,0 0,0002 0,0001 0,0000 0,0 0,0032 0,0009 0,0000
0,0000 0,0 0,0003 0,0006 0,0000 0,0 0,0002 0,0000 0,0 0,0001 0,0014 0,0000 0,0 9,0000 0,0 9,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000
0,0000 0,0 0,0040 0,0000 0,0004 0,0010 0,0001 0,0001 0,0001 0,0000 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0 0,0101 0,0000 0,0001 0,0000 0,0000 0,0000 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0 0,0000 0,0 0,0031 0,0000 0,0005 0,0000
0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0 0,0108 0,0000 0,0000 0,0002 0,0002 0,0002 0,0000 0,0001 0,0005 0,0013 0,0000 0,0000 0,0001 0,0000 0,0 0,0000,0 0

Біграми, що не перетинаються з пробілам:

			u	a	6	В	г	Д	e	ж	3	И	й	к	Л	М	н	0	п	р	С	Т	у	ф	х	ц	ч	Ш	щ	ъ	ы	ь	Э	ю	Я	ë
	(0,00	a58 0	,0027	0,0075	0,0148	0,0047	0,0072	0,0042	0,0020	0,0039	0,0113	0,0000	0,0103	0,0026	0,0056	0,0153	0,0120	0,0151	0,0040	0,0155	0,0068	0,0039	0,0007	0,0012	0,0003	0,0055	0,0008	0,0001	0,0000	0,0000	0,0000	0,0026	0,0000 (,0014 (9,0000
а		0,01	162 (0,0000	0,0006	0,0031	0,0008	0,0020	0,0009	0,0011	0,0037	0,0001	0,0008	0,0043	0,0090	0,0027	0,0046	0,0000	0,0008	0,0028	0,0035	0,0048	0,0001	0,0005	0,0010	0,0001	0,0007	0,0012	0,0002	0,0000	0,0000	0,0000	0,0000	0,0009	0,0024	0,0000
6		0,00	002 6	0,0010	0,0000	0,0001	0,0000	0,0000	0,0026	0,0000	0,0000	0,0007	0,0000	0,0002	0,0008	0,0000	0,0003	0,0022	0,0000	0,0011	0,0001	0,0000	0,0011	0,0000	0,0001	0,0000	0,0000	0,0000	0,0002	0,0001	0,0040	0,0000	0,0000	0,0000	0,0004	0,0000
В		0,00	061 (0,0053	0,0000	0,0000	0,0000	0,0002	0,0042	0,0000	0,0004	0,0030	0,0000	0,0001	0,0009	0,0001	0,0011	0,0070	0,0003	0,0006	0,0032	0,0002	0,0007	0,0000	0,0000	0,0000	0,0000	0,0011	0,0000	0,0000	0,0025	0,0002	0,0000	0,0000	0,0002	0,0000
г		0,00	007 (0,0009	0,0000	0,0000	0,0000	0,0009	0,0005	0,0000	0,0000	0,0007	0,0000	0,0001	0,0014	0,0000	0,0002	0,0089	0,0000	0,0011	0,0001	0,0000	0,0007	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Д											0,0000																									
е											0,0012																									
Ж											0,0000																									-
3											0,0001																									
И											0,0017																									_
И											0,0000																									
K.											0,0000																									-
,,,											0,0000 0,0000																									
											0,0000																									
											0,0008																									
_											0,0000																									
p											0,0000																									
c											0,0000																									
т		0,00	045 6	0,0056	0,0000	0,0024	0,0000	0,0002	0,0043	0,0000	0,0000	0,0032	0,0000	0,0004	0,0003	0,0000	0,0009	0,0138	0,0001	0,0026	0,0008	0,0001	0,0015	0,0000	0,0000	0,0001	0,0002	0,0000	0,0000	0,0000	0,0014	0,0054	0,0000	0,0001	0,0005	0,0000
У		0,00	062 6	0,0001	0,0005	0,0008	0,0012	0,0018	0,0001	0,0013	0,0007	0,0000	0,0001	0,0008	0,0013	0,0010	0,0003	0,0000	0,0006	0,0006	0,0012	0,0014	0,0000	0,0000	0,0004	0,0000	0,0008	0,0008	0,0002	0,0000	0,0000	0,0000	0,0000	0,0009	0,0001	0,0000
ф		0,00	002 6	0,0001	0,0000	0,0000	0,0000	0,0000	0,0001	0,0000	0,0000	0,0005	0,0000	0,0000	0,0001	0,0000	0,0000	0,0001	0,0000	0,0004	0,0000	0,0000	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
х		0,00	032 6	0,0009	0,0000	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0000	0,0000	0,0001	0,0000	0,0001	0,0022	0,0000	0,0001	0,0000	0,0000	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
ц		0,00	003 (0,0006	0,0000	0,0000	0,0000	0,0000	0,0009	0,0000	0,0000	0,0002	0,0000	0,0002	0,0000	0,0000	0,0000	0,0003	0,0000	0,0000	0,0000	0,0000	0,0004	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000
ч		0,00	002 (0,0020	0,0000	0,0000	0,0000	0,0000	0,0036	0,0000	0,0000	0,0014	0,0000	0,0003	0,0000	0,0000	0,0005	0,0001	0,0000	0,0000	0,0000	0,0037	0,0006	0,0000	0,0000	0,0000	0,0000	0,0001	0,0000	0,0000	0,0000	0,0002	0,0000	0,0000	0,0000	0,0000
Ш		0,00	001 (0,0014	0,0000	0,0000	0,0000	0,0000	0,0022	0,0000	0,0000	0,0016	0,0000	0,0004	0,0005	0,0000	0,0004	0,0003	0,0000	0,0000	0,0000	0,0001	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0003	0,0000	0,0000	0,0000	0,0000
щ		0,00	000 (0,0003	0,0000	0,0000	0,0000	0,0000	0,0012	0,0000	0,0000	0,0007	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
ъ		0,00	000 6	0,000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0000
ы		0,00	041 (0,000	0,0004	0,0011	0,0001	0,0001	0,0010	0,0000	0,0000	0,0000	0,0015	0,0002	0,0025	0,0014	0,0003	0,0000	0,0001	0,0003	0,0007	0,0005	0,0000	0,0000	0,0009	0,0000	0,0002	0,0005	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
ь											0,0001																									
Э											0,0000																									
Ю											0,0000																									
Я											0,0012																									
ë		0,00	000 6	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Біграми, що перетинаються без пробілів:

		a	б	В	г	Д	e	ж	3	И	й	К	л	М	н	0	п	р	С	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	Э	Ю	я	ë
'-																																		
а	0	,000	0,001	0,005	5 0,0015	0,0032	0,0016	0,0014	0,0048	0,0015	0,0009	0,0065	0,0109	0,0040	0,0077	0,0015	0,0028	0,0038	0,0060	0,0066	0,0006	0,0006	0,0013	0,0001	0,0014	0,0016	0,0002 (0,0000	0,0000	0,0000	0,0004	0,0010	0,0031	0,0000
6	0	,001	0,000	0,000	1 0,0000	0,0000	0,0025	0,0000	0,0000	0,0009	0,0000	0,0003	0,0009	0,0000	0,0004	0,0027	0,0000	0,0013	0,0001	0,0000	0,0014	0,0000	0,0001	0,0000	0,0000	0,0000	0,0002 (0,0001	0,0048	0,0001	0,0000	0,0000	0,0005	0,0000
В	0	,006	0,000	0,000	4 0,0004	0,0007	0,0053	0,0001	0,0007	0,0039	0,0000	0,0008	0,0013	0,0004	0,0019	0,0090	0,0011	0,0009	0,0046	0,0007	0,0010	0,0000	0,0001	0,0000	0,0002	0,0013	0,0000 (0,0000	0,0030	0,0002	0,0004	0,0000	0,0003	0,0000
r	0	,001:	0,000	0,000	1 0,0000	0,0012	0,0006	0,0000	0,0001	0,0009	0,0000	0,0001	0,0018	0,0000	0,0004	0,0107	0,0001	0,0014	0,0002	0,0000	0,0008	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000 (0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
д	0	,0049	0,000	1 0,001	2 0,0001	0,0001	0,0053	0,0000	0,0000	0,0030	0,0000	0,0004	0,0009	0,0001	0,0020	0,0043	0,0002	0,0026	0,0005	0,0004	0,0019	0,0000	0,0001	0,0002	0,0001	0,0001	0,0000 (0,0002	0,0007	0,0005	0,0000	0,0001	0,0005	0,0000
e	0	,000	0,0024	1 0,003	5 0,0047	0,0039	0,0023	0,0014	0,0021	0,0015	0,0029	0,0021	0,0077	0,0059	0,0100	0,0019	0,0031	0,0083	0,0063	0,0051	0,0006	0,0001	0,0014	0,0004	0,0017	0,0009	0,0008 (0,0000	0,0000	0,0000	0,0003	0,0004	0,0006	0,0000
ж	0	,001	0,000	L 0,000	0 0,0000	0,0008	0,0044	0,0000	0,0000	0,0015	0,0000	0,0001	0,0000	0,0000	0,0014	0,0001	0,0000	0,0000	0,0000	0,0000	0,0002	0,0000	0,0000	0,0000	0,0001	0,0000	0,0000 (0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
3	0	,0060	0,000	2 0,000	9 0,0005	0,0009	0,0004	0,0002	0,0001	0,0004	0,0000	0,0003	0,0003	0,0003	0,0020	0,0007	0,0002	0,0003	0,0003	0,0001	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000 (0,0000	0,0006	0,0010	0,0000	0,0001	0,0006	0,0000
и	0	,0004	0,0014	1 0,005	1 0,0012	0,0029	0,0033	0,0006	0,0027	0,0022	0,0015	0,0035	0,0060	0,0040	0,0057	0,0021	0,0026	0,0013	0,0055	0,0047	0,0006	0,0001	0,0020	0,0014	0,0020	0,0006	0,0001 (0,0000	0,0000	0,0000	0,0003	0,0004	0,0021	0,0000
й	0	,000	0,0004	1 0,000	7 0,0004	0,0008	0,0001	0,0002	0,0002	0,0007	0,0000	0,0008	0,0003	0,0004	0,0011	0,0006	0,0009	0,0004	0,0017	0,0006	0,0003	0,0001	0,0001	0,0001	0,0003	0,0002	0,0000 (0,0000	0,0000	0,0000	0,0001	0,0000	0,0001	0,0000
κ	0	,0084	0,000	0,000	7 0,0002	0,0003	0,0006	0,0002	0,0001	0,0032	0,0000	0,0004	0,0005	0,0002	0,0027	0,0103	0,0004	0,0020	0,0006	0,0008	0,0018	0,0000	0,0000	0,0000	0,0002	0,0000	0,0000 (0,0000	0,0000	0,0000	0,0001	0,0000	0,0001	0,0000
л	0	,008	0,0004	1 0,000	9 0,0004	0,0007	0,0050	0,0005	0,0002	0,0079	0,0000	0,0016	0,0002	0,0002	0,0011	0,0080	0,0011	0,0003	0,0024	0,0004	0,0016	0,0000	0,0000	0,0000	0,0007	0,0001	0,0000 (0,0000	0,0013	0,0042	0,0001	0,0012	0,0021	0,0000
м	0	,003	0,000	1 0,000	8 0,0004	0,0004	0,0035	0,0001	0,0002	0,0045	0,0000	0,0007	0,0003	0,0003	0,0020	0,0048	0,0011	0,0003	0,0010	0,0003	0,0032	0,0001	0,0001	0,0000	0,0005	0,0001	0,0000 (0,0000	0,0009	0,0000	0,0001	0,0000	0,0006	0,0000
н	0	,012	0,0004	1 0,000	6 0,0003	0,0016	0,0108	0,0001	0,0002	0,0091	0,0000	0,0005	0,0001	0,0001	0,0038	0,0105	0,0007	0,0002	0,0012	0,0008	0,0030	0,0000	0,0001	0,0006	0,0004	0,0001	0,0001 (0,0000	0,0036	0,0012	0,0001	0,0002	0,0035	0,0000
О	0	,000	0,005	0,011	6 0,0060	0,0062	0,0033	0,0024	0,0016	0,0024	0,0043	0,0032	0,0082	0,0068	0,0108	0,0025	0,0035	0,0079	0,0101	0,0093	0,0007	0,0005	0,0009	0,0001	0,0027	0,0015	0,0002 (0,0000	0,0000	0,0000	0,0004	0,0007	0,0011	0,0000
п	0	,001	0,000	0,000	0 0,0000	0,0000	0,0030	0,0000	0,0000	0,0009	0,0000	0,0001	0,0008	0,0000	0,0001	0,0101	0,0001	0,0071	0,0000	0,0000	0,0008	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000 (0,0000	0,0003	0,0008	0,0000	0,0000	0,0004	0,0000
p	0	,009	0,000	2 0,000	5 0,0004	0,0005	0,0073	0,0004	0,0001	0,0060	0,0000	0,0004	0,0001	0,0003	0,0009	0,0078	0,0003	0,0001	0,0005	0,0008	0,0033	0,0000	0,0001	0,0001	0,0001	0,0003	0,0001 (0,0000	0,0019	0,0011	0,0000	0,0002	0,0012	0,0000
c	0	,001	0,000	2 0,002	4 0,0001	0,0005	0,0044	0,0001	0,0001	0,0018	0,0000	0,0055	0,0029	0,0013	0,0014	0,0033	0,0023	0,0004	0,0012	0,0132	0,0011	0,0000	0,0002	0,0000	0,0005	0,0001	0,0000 (0,0000	0,0004	0,0035	0,0001	0,0001	0,0036	0,0000
т	0	,006	0,000	0,003	3 0,0002	0,0004	0,0052	0,0001	0,0001	0,0043	0,0000	0,0007	0,0004	0,0002	0,0016	0,0169	0,0005	0,0033	0,0015	0,0004	0,0019	0,0000	0,0001	0,0002	0,0004	0,0000	0,0000 (0,0000	0,0017	0,0066	0,0001	0,0001	0,0007	0,0000
У	0	,000	0,000	0,001	6 0,0016	0,0024	0,0003	0,0017	0,0010	0,0007	0,0001	0,0015	0,0017	0,0015	0,0012	0,0005	0,0014	0,0009	0,0021	0,0019	0,0001	0,0000	0,0005	0,0000	0,0014	0,0009	0,0003 (0,0000	0,0000	0,0000	0,0001	0,0011	0,0002	0,0000
ф	0	,000	0,000	0,000	0 0,0000	0,0000	0,0002	0,0000	0,0000	0,0007	0,0000	0,0000	0,0001	0,0000	0,0000	0,0001	0,0000	0,0005	0,0000	0,0000	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000 (0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
×	0	,001	0,000	2 0,000	5 0,0001	0,0002	0,0001	0,0000	0,0001	0,0005	0,0000	0,0002	0,0002	0,0002	0,0004	0,0030	0,0004	0,0002	0,0005	0,0001	0,0002	0,0000	0,0000	0,0000	0,0001	0,0000	0,0000 (0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
ц	0	,000	0,000	0,000	1 0,0000	0,0000	0,0011	0,0000	0,0000	0,0003	0,0000	0,0002	0,0000	0,0000	0,0000	0,0004	0,0000	0,0000	0,0000	0,0000	0,0005	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000 (0,0000	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000
ч	0	,0024	0,000	0,000	0 0,0000	0,0000	0,0037	0,0000	0,0000	0,0017	0,0000	0,0003	0,0000	0,0000	0,0006	0,0001	0,0000	0,0000	0,0000	0,0044	0,0007	0,0000	0,0000	0,0000	0,0000	0,0001	0,0000 (0,0000	0,0000	0,0002	0,0000	0,0000	0,0000	0,0000
ш	0	,001	0,000	0,000	0 0,0000	0,0000	0,0026	0,0000	0,0000	0,0020	0,0000	0,0005	0,0006	0,0000	0,0005	0,0004	0,0001	0,0000	0,0000	0,0001	0,0004	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000 (0,0000	0,0000	0,0003	0,0000	0,0000	0,0000	0,0000
щ	0	,0004	0,000	0,000	0 0,0000	0,0000	0,0015	0,0000	0,0000	0,0008	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000 (0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
ъ	0	,0000	0,000	0,000	0 0,0000	0,0000	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000 (0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0000
ы	0	,000	0,000	3 0,001	7 0,0002	0,0003	0,0014	0,0001	0,0002	0,0005	0,0018	0,0005	0,0031	0,0019	0,0009	0,0003	0,0006	0,0004	0,0013	0,0009	0,0001	0,0000	0,0012	0,0000	0,0003	0,0006	0,0000 (0,0000	0,0000	0,0000	0,0001	0,0000	0,0001	0,0000
ь	0	,000	0,000	0,001	3 0,0003	0,0005	0,0020	0,0001	0,0004	0,0011	0,0000	0,0020	0,0002	0,0007	0,0024	0,0009	0,0010	0,0002	0,0021	0,0006	0,0002	0,0000	0,0001	0,0001	0,0005	0,0005	0,0000 (0,0000	0,0000	0,0000	0,0002	0,0005	0,0006	0,0000
э	0	,0000	0,000	0,000	0 0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0028	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000 (0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Ю	0	,000	0,000	7 0,000	3 0,0001	0,0005	0,0001	0,0001	0,0001	0,0003	0,0000	0,0003	0,0001	0,0002	0,0003	0,0002	0,0003	0,0002	0,0004	0,0006	0,0001	0,0000	0,0000	0,0000	0,0002	0,0001	0,0005	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
я	0	,000	0,000	0,001	6 0,0005	0,0013	0,0005	0,0007	0,0019	0,0011	0,0000	0,0011	0,0012	0,0008	0,0023	0,0009	0,0011	0,0003	0,0021	0,0022	0,0003	0,0001	0,0003	0,0001	0,0005	0,0001	0,0002 (0,0000	0,0000	0,0000	0,0002	0,0001	0,0003	0,0000
ë	0	,0000	0,000	0,000	0 0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Біграми, що не перетинаються без пробілів:

0,0003 0,0015 0,0055 0,0014 0,0032 0,0017 0,0014 0,0049 0,0015 0,0009 0,0066 0,0109 0,0066 0,0014 0,0028 0,0018 0,0006 0,0006 0,0006 0,0001 0,0002 0,0015 0,0016 0,0003 0,0000 0,0 0,0012 0,0000 0,0001 0,0000 0,0 0,0066 0,0003 0,0004 0,0004 0,0006 0,0053 0,0001 0,0007 0,0039 0,0000 0,0008 0,0013 0,0004 0,0019 0,0089 0,0010 0,0007 0,0010 0,0000 0,0001 0,0000 0,0002 0,0013 0,0000 0,002 0,0002 0,0003 0,0000 0,00 0,0011 0,0000 0,0001 0,0000 0,0 0,0050 0,0001 0,0012 0,0001 0,0001 0,0001 0,0001 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0001 0,0002 0,0001 0,0000 0,0001
0,0001 0,0 0,0002 0,0024 0,0035 0,0047 0,0039 0,0023 0,0014 0,0021 0,0015 0,0029 0,0021 0,007 0,0058 0,0101 0,0019 0,0083 0,0063 0,0063 0,0006 0,0001 0,0014 0,0004 0,0017 0,0009 0,0008 0,0000 0,0000 0,0000 0,0000 0,0004 0,0006 0,0000 0,0015 0,0001 0,0000 0,0 0,0060 0,0002 0,0009 0,0005 0,0009 0,0004 0,0002 0,0001 0,0004 0,0000 0,0003 0,0002 0,0004 0,0002 0,0007 0,0002 0,0003 0,0001 0,0000 0,0 0,0003 0,0014 0,0052 0,0012 0,0029 0,0033 0,0006 0,0026 0,0022 0,0015 0,0035 0,0059 0,0039 0,0057 0,0021 0,0027 0,0013 0,0055 0,0047 0,0006 0,0001 0,0020 0,0014 0,0020 0,0006 0,0002 0,0000 0,0000 0,0000 0,0000 0,0004 0,0001 0,0000 0,0002 0,0004 0,0007 0,0004 0,0008 0,0001 0,0002 0,0002 0,0000 0,0000 0,0008 0,0003 0,0004 0,0012 0,0006 0,0009 0,0004 0,0017 0,0006 0,0001 0,0001 0,0001 0,0001 0,0001 0,0000 0,0000 0,0000 0,0000 0,0001 0,0001 0,0000 0,0001 0,0000 0,0083 0,0005 0,0007 0,0002 0,0003 0,0007 0,0002 0,0000 0,0 0,0082 0,0004 0,0009 0,0004 0,0007 0,0050 0,0005 0,0005 0,0005 0,0000 0,0015 0,0002 0,0002 0,0012 0,0012 0,0010 0,0 0,0034 0,0004 0,0009 0,0004 0,0004 0,0004 0,0001 0,0002 0,0046 0,0000 0,0007 0,0003 0,0003 0,0003 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0000 0,0000 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001
0,0001 0,0 0,0123 0,0004 0,0006 0,0003 0,0016 0,0107 0,0001 0,0002 0,0000 0,0000 0,0000 0,0001 0,0 0,0002 0,0053 0,0115 0,0061 0,0062 0,0033 0,0024 0,0016 0,0024 0,0016 0,0024 0,0082 0,0082 0,0082 0,0087 0,0108 0,0025 0,0036 0,0093 0,0007 0,0005 0,0005 0,0009 0,0001 0,0027 0,0015 0,0002 0,0000 0,0000 0,0000 0,0000 0,0007 0,0011 0,0000 0,0015 0,0000 0,0 0,0095 0,0002 0,0004 0,0005 0,0004 0,0005 0,0004 0,0001 0,0000 0,0004 0,0001 0,0000 0,0004 0,0001 0,0003 0,0010 0,0003 0,0001 0,0000 0,0001 0,0000 0,0001 0,0001 0,0000 0,0010 0,0001 0,0001 0,0000 0,0000 0,0000 0,0 0,0018 0,0002 0,0024 0,0001 0,0005 0,0043 0,0001 0,0001 0,0001 0,0001 0,0001 0,0001 0,0005 0,0025 0,0021 0,0001 0,0002 0,0002 0,0001 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0001 0,0 0,0067 0,0003 0,0003 0,0002 0,0004 0,0052 0,0001 0,0001 0,0004 0,0000 0,0007 0,0004 0,0002 0,0016 0,0170 0,0005 0,0032 0,0015 0,0004 0,0010 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0018 0,0010 0,0007 0,0007 0,0007 0,0007 0,0000 0,0002 0,0009 0,0016 0,0016 0,0013 0,0003 0,0007 0,0010 0,0008 0,0001 0,0015 0,0017 0,0015 0,0015 0,0012 0,0005 0,0014 0,0009 0,0001 0,0000 0,0000 0,0001 0,0010
0,0010 0,0001 0,0000 0,0000 0,0001 0,0000 0,0001 0,0001 0,0000 0,0000 0,0000 0,0000 0,0001 0,0001 0,0002 0,0000 0,0002 0,0000 0,0 0,0011 0,0002 0,0001 0,0001 0,0002 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0 0,0007 0,0000 0,0001 0,0000 0,0 0,0025 0,0000 0,0 0,0017 0,0000
0,0000 0,0 0,0003 0,0000 0,000 0,000 0,000 0,0000 0,0000 0,0000 0,000 0,0 0,0001 0,0008 0,0017 0,0003 0,0003 0,0004 0,0001 0,0000 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0001 0,0000 0,0 0,0007 0,0006 0,0013 0,0003 0,0005 0,0019 0,0001 0,0004 0,0011 0,0000 0,0002 0,0007 0,0004 0,0010 0,0002 0,0001 0,0006 0,0003 0,0000 0,0001 0,0001 0,0005 0,0000 0,0000 0,0000 0,0000 0,0000 0,0000 0,0007 0,0007 0,0000 0,0 0,0001 0,0007 0,0003 0,0001 0,0005 0,0001 0,0001 0,0001 0,0001 0,0001 0,0000
0,0000 0,0 0,0004 0,0005 0,0016 0,0005 0,0013 0,0005 0,0007 0,0018 0,0011 0,0000 0,0011 0,0002 0,0008 0,0023 0,0009 0,0001 0,0002 0,0001 0,0003 0,0001 0,0005 0,0001 0,0002 0,0000 0,0000 0,0000 0,0002 0,0001 0,0003 0,0000 0,000 0,00

Ентропія ймовірнісного ансамблю:

$$H(Z) = -\sum_{i=1}^{n} p_i \log p_i$$

де п-кількість літер у тексті, рі-це частота і-того символу.

Надлишковість джерела відкритого тексту:

$$R = 1 - \frac{H_{\infty}}{H_0}$$

Ентропія та надмірність з пробілами:

Ентропія та надмірність:

H1: 4.3698544396

Н2 (Перетин): 3.5758799715

Н2 (Без перетину): 2.2878734706

R1: 0.1410542787

R2 (Перетин): 0.2971191961

R2 (Без перетину): 0.5502918563

Ентропія та надмірність без пробілів:

Ентропія та надмірність:

H1: 4.4667340453

Н2 (Перетин): 3.8277814147

Н2 (Без перетину): 2.4136655851

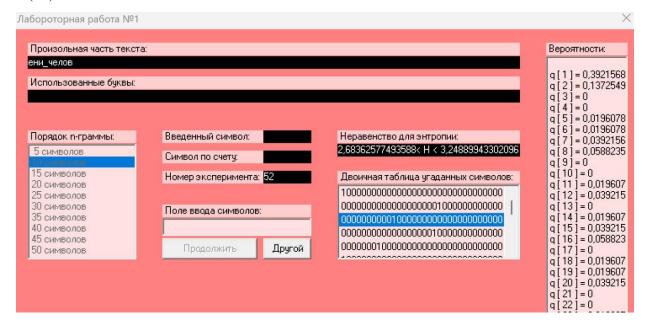
R1: 0.1145152540

R2 (Перетин): 0.2411811361

R2 (Без перетину): 0.5215152647

2. За допомогою програми CoolPinkProgram оцінити значення H(10), H(20), H(30).

H(10):



2,6836<H<3,2489

$$1. R = 1 - \frac{2,6836}{5} = 0,46328$$

2.
$$R = 1 - \frac{3,2489}{5} = 0,35022$$

0,35022<R<0,46328

H(20):

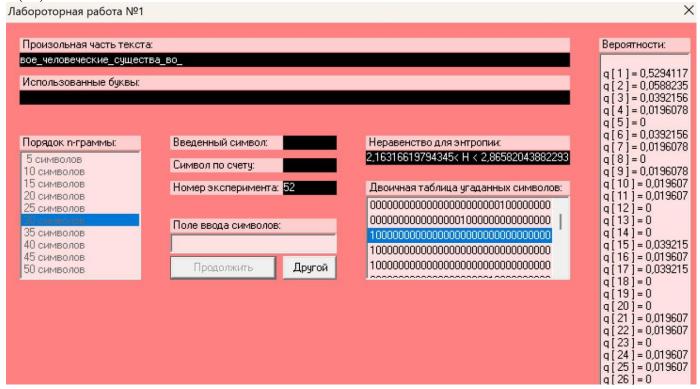


1,8669<H<2,4883

$$1. R = 1 - \frac{1,8669}{5} = 0,62662$$

2.
$$R = 1 - \frac{2,4883}{5} = 0,50234$$

H(30):



2,1637<H<2,8658

1.
$$R = 1 - \frac{2,1637}{5} = 0,56726$$

$$2. R = 1 - \frac{2,8658}{5} = 0,42684$$

0,42684<R<0,56726

Висновок:

У ході виконання роботи було закріплено поняття ентропії на символ джерела та надлишковості, а також досліджено різні моделі джерела відкритого тексту. Було реалізовано програмний підхід для підрахунку частот літер і біграм, що дало змогу визначити значення ентропії H_1 та H_2 для тексту з пробілами та без них. Отримані результати показали, що вилучення пробілів впливає на значення ентропії та надлишковості, оскільки змінюється розподіл імовірностей символів.

Крім того, за допомогою програми CoolPinkProgram було оцінено значення H(10), H(20), H(30).