

Національний Технічний Університет України  
“Київський Політехнічний Інститут”  
Фізико-Технічний Інститут

# СИМЕТРИЧНА КРИПТОГРАФІЯ

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1:

Експериментальна оцінка ентропії на символ джерела  
відкритого тексту

Виконав студент 3-го курсу  
групи ФІ-14  
Геращенко Володимир

Київ 2024

# 1 Мета роботи

Засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкрито го тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.

## 2 Хід роботи

### 2.1 Робота з текстом

Код для роботи з текстом і перетворення його у зручний для аналізу вигляд:

```
1 public static class Reformator
2 {
3     const string alphabet = "АБВГДЕЁЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯабвгдеёжзийклмнопрсту
4     фхцчшщъыьэюя ";
5     private static string ReformatString(string s, bool withSpaces)
6     {
7         string temp = "";
8         foreach (char c in s)
9         {
10             temp += alphabet.Contains(c) ? c : " ";
11         }
12         temp = Regex.Replace(temp, @"\s+", withSpaces ? " " : "");
13         temp = temp.ToLower();
14         return temp;
15     }
16     public static void ReformatFile(string inputFile, string outputFile, bool
17     withSpaces = true)
18     {
19         var reader = new StreamReader(inputFile, Encoding.UTF8);
20         using StreamWriter sw = new StreamWriter(outputFile);
21         while (!reader.EndOfStream)
22         {
23             string? line = reader.ReadLine();
24             if (line == null) break;
25             sw.Write(ReformatString(line, withSpaces));
26         }
27     }
28 }
```

Самий зручний формат для роботи з текстом є .txt, тому всі файли будуть саме в цьому форматі. Але на цьому етапі вже виникла проблема: клас для роботи з файлами ніяк не міг зчитати текст з файлу - то зчитував одні пробіли, то виводив незрозумілі літери. Зміна енкодеру ніяк не допомогала, але рішення виявилось доволі простим: треба було змінити енкодінг не в коді, а у самому файлі - після цього все почало працювати добре.

### 2.2 Підрахунок кількості літер і біграм

Далі розглянемо методи для підрахунку кількості літер і біграм у тексті:

1. Підрахунок кількості літер у тексті:

```
1 public ulong[] LetterFrequencies()
2 {
3     ulong[] frequencies = new ulong[alphabet.Length - (WithSpaces ? 0 : 1)];
4     if (File.Exists(inputFile))
5     {
6         string buffer = File.ReadAllText(inputFile);
7         foreach (var letter in buffer)
8         {
9             frequencies[alphabet.IndexOf(letter)] += 1;
10        }
11    }
12    return frequencies;
13 }
14 }
```

Я пішов самим простим шляхом - просто зчитав з файлу одну велику строку у буфер - тому для **дуже** великих текстів буде важко його запустити.

## 2. Підрахунок ймовірності зустріти кожну літеру:

```
1 public float[] LetterProbabilities()
2 {
3     float[] probabilities = new float[alphabet.Length - (WithSpaces ? 0 : 1)];
4     ulong[] frequencies = LetterFrequencies();
5
6     var textSize = TextSize();
7
8     for(int i = 0; i < probabilities.Length; i++)
9     {
10         probabilities[i] = (float)frequencies[i]/textSize;
11     }
12     return probabilities;
13 }
14
```

## 3. Підрахунок загальної кількості літер у тексті:

```
1 public ulong TextSize()
2 {
3     string s = "";
4     if (File.Exists(inputFile))
5     {
6         s = File.ReadAllText(inputFile);
7     }
8     return (ulong)s.Length;
9 }
10
```

Також код для підрахунку кількості і ймовірності біграм:

### 1. Підрахунок кількості біграм у тексті:

```
1 public (ulong[,] frequencies, ulong size) BigramFrequencies(int step = 1)
2 {
3     ulong[,] frequencies = new ulong[alphabet.Length - (WithSpaces ? 0 : 1),
4     alphabet.Length - (WithSpaces ? 0 : 1)];
5     ulong size = 0;
6     if (File.Exists(inputFile))
7     {
8         string s = File.ReadAllText(inputFile);
9
10        for (int i = 0; i < s.Length - 1; i += step)
11        {
12            frequencies[alphabet.IndexOf(s[i]), alphabet.IndexOf(s[i + 1])] += 1;
13            size++;
14        }
15    }
16    return (frequencies, size);
17 }
```

Спочатку я думав для підрахунку ймовірності брати TextSize - 1, але я зрозумів, що такий підхід не спрацює для біграм з кроком 2, тому я параметризував цю функцію, і повертаю ще кількість усіх біграм одразу.

### 2. Підрахунок ймовірності появи кожної біграми:

```
1 public float[,] BigramProbabilities(int step = 1)
2 {
3     float[,] probabilities = new float[alphabet.Length - (WithSpaces ? 0 : 1),
4     alphabet.Length - (WithSpaces ? 0 : 1)];
5     var frequencies = BigramFrequencies(step);
6     for (int i = 0; i < alphabet.Length - (WithSpaces ? 0 : 1); i++)
7     {
8         for (int j = 0; j < alphabet.Length - (WithSpaces ? 0 : 1); j++)
9         {
10             probabilities[i, j] = (float)frequencies.frequencies[i, j] / (float)
11             frequencies.size;
12         }
13     }
14     return probabilities;
15 }
```

## 2.3 Підрахунок ентропії

Підрахунок ентропії  $H_1$ :

```
1 public float LetterEntropy()  
2 {  
3     float result = 0;  
4     var probabilities = LetterProbabilities();  
5     for (int i = 0; i < alphabet.Length - (WithSpaces ? 0 : 1); i++)  
6     {  
7         var log = probabilities[i] != 0 ? Math.Log(probabilities[i], 2) : 0;  
8         result += (float)(log * probabilities[i]);  
9     }  
10    return -result;  
11 }
```

Підрахунок ентропії  $H_2$ :

```
1 public float BigramEntropy(int step = 1)  
2 {  
3     float result = 0;  
4     var probabilities = BigramProbabilities(step);  
5     for (int i = 0; i < alphabet.Length - (WithSpaces ? 0 : 1); i++)  
6     {  
7         for(int j = 0; j < alphabet.Length - (WithSpaces ? 0 : 1); j++)  
8         {  
9             var log = probabilities[i,j] != 0 ? Math.Log(probabilities[i,j], 2) : 0;  
10            result += (float)(log * probabilities[i,j]);  
11        }  
12    }  
13    return -result/2;  
14 }  
15
```

Так як функція логарифму в C# повертає  $-\infty$  при значенні  $= 0$ , то була певна проблема з невизначеностями, і значення в результаті було NaN. Прийшлося самому розглядати випадок при нульовій ймовірності.

## 3 Результати роботи

### 3.1 Підрахунок кількості літер і біграм

Нижче можемо побачити таблиці частот для тексту з пробілами, та без пробілів. Сторінками нижче ідуть великі таблиці ймовірностей для біграм у вигляді у відсотках, тому що так поміститься набагато більше інформації. Також розглянуто випадки, коли для підрахунку біграм ми рухаємося з кроком 1 і з кроком 2.

Літера	Кількість	Ймовірність
—	605032	0.1666
о	349938	0.0964
е	256299	0.0706
а	247726	0.0682
и	204783	0.0564
н	201791	0.0556
т	179227	0.0494
с	160514	0.0442
л	152184	0.0419
в	141509	0.0390
р	128159	0.0353
к	102810	0.0283
д	91629	0.0252
м	87872	0.0242
у	82543	0.0227
п	77502	0.0213
я	66064	0.0182
ь	59161	0.0163
ы	57627	0.0159
г	57603	0.0159
б	52430	0.0144
з	50903	0.0140
ч	47066	0.0130
ж	33341	0.0092
й	32456	0.0089
ш	26910	0.0074
х	24700	0.0068
ю	18633	0.0051
э	9915	0.0027
ц	9885	0.0027
щ	8799	0.0024
ф	5052	0.0014
ъ	1151	0.0003
ё	488	0.0001

Табл. 1: Таблиця частот для тексту з пробілами

Літера	Кількість	Ймовірність
о	349938	0.1156
е	256299	0.0847
а	247726	0.0818
и	204783	0.0677
н	201791	0.0667
т	179227	0.0592
с	160514	0.0530
л	152184	0.0503
в	141509	0.0468
р	128159	0.0423
к	102810	0.0340
д	91629	0.0303
м	87872	0.0290
у	82543	0.0273
п	77502	0.0256
я	66064	0.0218
ь	59161	0.0195
ы	57627	0.0190
г	57603	0.0190
б	52430	0.0173
з	50903	0.0168
ч	47066	0.0156
ж	33341	0.0110
й	32456	0.0107
ш	26910	0.0089
х	24700	0.0082
ю	18633	0.0062
э	9915	0.0033
ц	9885	0.0033
щ	8799	0.0029
ф	5052	0.0017
ъ	1151	0.0004
ё	488	0.0002

Табл. 2: Таблиця частот для тексту без пробілів

	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	0.00%	0.06%	0.32%	0.07%	0.20%	0.10%	0.00%	0.12%	0.40%	0.01%	0.06%	0.40%	0.91%	0.25%	0.45%	0.00%	0.08%	0.26%	0.34%	0.43%	0.01%	0.04%	0.09%	0.01%	0.08%	0.09%	0.02%	0.00%	0.00%	0.00%	0.00%	0.08%	0.23%	1.72%
б	0.09%	0.00%	0.01%	0.00%	0.00%	0.20%	0.00%	0.00%	0.00%	0.07%	0.00%	0.02%	0.07%	0.00%	0.03%	0.21%	0.00%	0.11%	0.01%	0.00%	0.10%	0.00%	0.01%	0.00%	0.00%	0.00%	0.02%	0.01%	0.40%	0.00%	0.00%	0.00%	0.05%	0.03%
в	0.54%	0.00%	0.00%	0.00%	0.03%	0.42%	0.00%	0.00%	0.05%	0.38%	0.00%	0.01%	0.09%	0.01%	0.12%	0.72%	0.02%	0.08%	0.32%	0.02%	0.07%	0.00%	0.00%	0.00%	0.01%	0.12%	0.00%	0.00%	0.24%	0.02%	0.00%	0.00%	0.02%	0.58%
г	0.09%	0.00%	0.00%	0.00%	0.11%	0.05%	0.00%	0.00%	0.00%	0.07%	0.00%	0.01%	0.14%	0.00%	0.02%	0.85%	0.00%	0.09%	0.01%	0.00%	0.06%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.08%
д	0.42%	0.00%	0.09%	0.00%	0.00%	0.44%	0.00%	0.00%	0.00%	0.23%	0.00%	0.02%	0.08%	0.01%	0.16%	0.36%	0.01%	0.16%	0.04%	0.02%	0.17%	0.00%	0.01%	0.02%	0.00%	0.01%	0.00%	0.01%	0.06%	0.06%	0.00%	0.00%	0.04%	0.11%
е	0.00%	0.12%	0.19%	0.35%	0.25%	0.18%	0.00%	0.09%	0.11%	0.01%	0.24%	0.13%	0.58%	0.39%	0.72%	0.04%	0.11%	0.59%	0.37%	0.39%	0.01%	0.00%	0.09%	0.02%	0.09%	0.08%	0.06%	0.00%	0.00%	0.00%	0.00%	0.03%	0.03%	1.78%
ё	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	
ж	0.14%	0.00%	0.00%	0.00%	0.07%	0.39%	0.00%	0.00%	0.00%	0.14%	0.00%	0.01%	0.00%	0.00%	0.11%	0.00%	0.00%	0.00%	0.01%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	
з	0.52%	0.02%	0.08%	0.04%	0.07%	0.02%	0.00%	0.01%	0.00%	0.03%	0.00%	0.01%	0.02%	0.03%	0.17%	0.06%	0.00%	0.02%	0.01%	0.00%	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.05%	0.04%	0.00%	0.00%	0.05%	0.12%
и	0.01%	0.04%	0.25%	0.04%	0.15%	0.26%	0.00%	0.04%	0.18%	0.07%	0.13%	0.17%	0.45%	0.28%	0.34%	0.01%	0.02%	0.05%	0.25%	0.34%	0.00%	0.00%	0.16%	0.10%	0.15%	0.04%	0.01%	0.00%	0.00%	0.00%	0.00%	0.03%	0.18%	1.87%
й	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.03%	0.00%	0.00%	0.00%	0.06%	0.02%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.72%	
к	0.73%	0.00%	0.03%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.26%	0.00%	0.00%	0.03%	0.00%	0.11%	0.79%	0.00%	0.14%	0.05%	0.04%	0.14%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.46%	
л	0.69%	0.00%	0.00%	0.01%	0.02%	0.46%	0.00%	0.04%	0.00%	0.58%	0.00%	0.04%	0.02%	0.00%	0.03%	0.59%	0.01%	0.00%	0.11%	0.00%	0.11%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.00%	0.10%	0.36%	0.00%	0.10%	0.18%	0.72%
м	0.25%	0.00%	0.00%	0.00%	0.00%	0.28%	0.00%	0.00%	0.00%	0.29%	0.00%	0.01%	0.01%	0.00%	0.11%	0.33%	0.01%	0.01%	0.01%	0.00%	0.25%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.09%	0.00%	0.00%	0.00%	0.04%	0.72%
н	1.05%	0.00%	0.00%	0.01%	0.07%	0.92%	0.00%	0.00%	0.00%	0.76%	0.00%	0.02%	0.00%	0.00%	0.27%	0.90%	0.00%	0.01%	0.07%	0.04%	0.23%	0.00%	0.00%	0.05%	0.02%	0.00%	0.01%	0.00%	0.30%	0.11%	0.00%	0.01%	0.21%	0.48%
о	0.00%	0.33%	0.74%	0.46%	0.45%	0.24%	0.00%	0.18%	0.10%	0.08%	0.33%	0.16%	0.61%	0.47%	0.73%	0.02%	0.12%	0.58%	0.63%	0.65%	0.01%	0.02%	0.04%	0.01%	0.14%	0.11%	0.01%	0.00%	0.00%	0.00%	0.06%	0.06%	2.31%	
п	0.12%	0.00%	0.00%	0.00%	0.00%	0.22%	0.00%	0.00%	0.00%	0.07%	0.00%	0.01%	0.06%	0.00%	0.01%	0.84%	0.00%	0.62%	0.00%	0.00%	0.06%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.03%	0.05%	0.00%	0.00%	0.03%	0.00%
р	0.75%	0.02%	0.03%	0.03%	0.03%	0.55%	0.00%	0.03%	0.00%	0.47%	0.00%	0.04%	0.01%	0.02%	0.06%	0.66%	0.01%	0.00%	0.02%	0.05%	0.26%	0.00%	0.01%	0.00%	0.01%	0.03%	0.00%	0.00%	0.14%	0.09%	0.00%	0.01%	0.10%	0.10%
с	0.16%	0.00%	0.18%	0.00%	0.03%	0.37%	0.01%	0.00%	0.00%	0.14%	0.00%	0.48%	0.25%	0.09%	0.08%	0.26%	0.17%	0.02%	0.08%	1.06%	0.07%	0.00%	0.02%	0.00%	0.04%	0.01%	0.00%	0.03%	0.28%	0.00%	0.01%	0.30%	0.28%	
т	0.51%	0.00%	0.25%	0.00%	0.01%	0.46%	0.00%	0.00%	0.00%	0.36%	0.00%	0.04%	0.02%	0.00%	0.10%	1.44%	0.01%	0.25%	0.09%	0.01%	0.13%	0.00%	0.00%	0.01%	0.02%	0.00%	0.00%	0.00%	0.14%	0.57%	0.00%	0.00%	0.05%	0.44%
у	0.00%	0.05%	0.10%	0.11%	0.16%	0.02%	0.00%	0.15%	0.06%	0.00%	0.01%	0.08%	0.12%	0.10%	0.03%	0.00%	0.06%	0.05%	0.12%	0.12%	0.00%	0.00%	0.03%	0.00%	0.08%	0.07%	0.02%	0.00%	0.00%	0.00%	0.00%	0.09%	0.01%	0.62%
ф	0.01%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.04%	0.00%	0.00%	0.01%	0.00%	0.00%	0.01%	0.00%	0.04%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	
х	0.08%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.01%	0.01%	0.01%	0.21%	0.00%	0.01%	0.01%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.31%	
ц	0.05%	0.00%	0.00%	0.00%	0.00%	0.08%	0.00%	0.00%	0.00%	0.02%	0.00%	0.01%	0.00%	0.00%	0.00%	0.03%	0.00%	0.00%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.02%	
ч	0.21%	0.00%	0.00%	0.00%	0.00%	0.32%	0.00%	0.00%	0.00%	0.13%	0.00%	0.03%	0.00%	0.00%	0.05%	0.01%	0.00%	0.00%	0.00%	0.40%	0.07%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.04%	
ш	0.12%	0.00%	0.00%	0.00%	0.00%	0.23%	0.00%	0.00%	0.00%	0.17%	0.00%	0.03%	0.05%	0.00%	0.04%	0.03%	0.00%	0.00%	0.00%	0.01%	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.00%	0.00%	0.01%	
щ	0.03%	0.00%	0.00%	0.00%	0.00%	0.12%	0.00%	0.00%	0.00%	0.07%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
ъ	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.00%	
ы	0.00%	0.04%	0.10%	0.01%	0.01%	0.10%	0.00%	0.00%	0.01%	0.00%	0.13%	0.02%	0.25%	0.13%	0.03%	0.00%	0.01%	0.03%	0.08%	0.07%	0.00%	0.00%	0.09%	0.00%	0.02%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.42%	
ь	0.00%	0.01%	0.00%	0.00%	0.00%	0.09%	0.00%	0.00%	0.01%	0.02%	0.00%	0.11%	0.00%	0.02%	0.11%	0.00%	0.00%	0.08%	0.01%	0.00%	0.00%	0.00%	0.00%	0.01%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.04%	0.04%	1.00%	
э	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.25%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
ю	0.00%	0.05%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.04%	0.00%	0.00%	0.00%	0.01%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.31%	
я	0.00%	0.00%	0.03%	0.02%	0.06%	0.01%	0.00%	0.04%	0.06%	0.00%	0.01%	0.01%	0.08%	0.03%	0.06%	0.00%	0.00%	0.01%	0.08%	0.14%	0.00%	0.00%	0.02%	0.00%	0.01%	0.00%	0.02%	0.00%	0.00%	0.00%	0.01%	0.02%	1.09%	
—	0.27%	0.69%	1.47%	0.40%	0.74%	0.46%	0.00%	0.22%	0.40%	1.12%	0.00%	0.95%	0.31%	0.54%	1.58%	1.27%	1.49%	0.40%	1.57%	0.80%	0.39%	0.07%	0.13%	0.03%	0.59%	0.08%	0.01%	0.00%	0.00%	0.27%	0.00%	0.17%	0.25%	

Табл. 3: Таблица частот біграм для тексту з пробілами (біграми 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	0.04%	0.15%	0.56%	0.13%	0.34%	0.19%	0.00%	0.16%	0.52%	0.16%	0.07%	0.62%	1.13%	0.38%	0.75%	0.17%	0.27%	0.36%	0.61%	0.61%	0.06%	0.05%	0.12%	0.01%	0.15%	0.11%	0.03%	0.00%	0.00%	0.00%	0.04%	0.10%	0.31%
б	0.11%	0.00%	0.01%	0.00%	0.00%	0.24%	0.00%	0.00%	0.00%	0.09%	0.00%	0.03%	0.09%	0.01%	0.04%	0.26%	0.00%	0.13%	0.01%	0.00%	0.12%	0.00%	0.01%	0.00%	0.00%	0.00%	0.02%	0.01%	0.48%	0.00%	0.01%	0.00%	0.06%
в	0.66%	0.02%	0.04%	0.04%	0.08%	0.53%	0.00%	0.01%	0.07%	0.48%	0.00%	0.08%	0.13%	0.04%	0.19%	0.91%	0.10%	0.12%	0.46%	0.08%	0.10%	0.00%	0.01%	0.00%	0.02%	0.15%	0.00%	0.00%	0.29%	0.02%	0.03%	0.00%	0.03%
г	0.11%	0.00%	0.01%	0.00%	0.14%	0.06%	0.00%	0.00%	0.00%	0.09%	0.00%	0.02%	0.17%	0.00%	0.03%	1.03%	0.01%	0.11%	0.02%	0.00%	0.08%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
д	0.51%	0.01%	0.12%	0.01%	0.01%	0.53%	0.00%	0.00%	0.00%	0.29%	0.00%	0.03%	0.10%	0.01%	0.21%	0.44%	0.02%	0.20%	0.06%	0.03%	0.20%	0.00%	0.01%	0.02%	0.01%	0.01%	0.00%	0.02%	0.07%	0.07%	0.00%	0.00%	0.05%
е	0.03%	0.24%	0.43%	0.47%	0.40%	0.26%	0.00%	0.13%	0.21%	0.16%	0.29%	0.26%	0.75%	0.59%	1.04%	0.20%	0.34%	0.76%	0.66%	0.57%	0.07%	0.01%	0.13%	0.02%	0.16%	0.10%	0.07%	0.00%	0.00%	0.00%	0.03%	0.04%	0.06%
ё	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ж	0.17%	0.01%	0.00%	0.00%	0.09%	0.47%	0.00%	0.00%	0.00%	0.17%	0.00%	0.01%	0.00%	0.00%	0.13%	0.01%	0.00%	0.00%	0.01%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
з	0.63%	0.02%	0.11%	0.06%	0.09%	0.03%	0.00%	0.02%	0.01%	0.04%	0.00%	0.03%	0.03%	0.04%	0.22%	0.07%	0.02%	0.03%	0.03%	0.01%	0.03%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.06%	0.05%	0.00%	0.00%	0.06%
и	0.03%	0.14%	0.53%	0.11%	0.28%	0.37%	0.00%	0.06%	0.28%	0.23%	0.15%	0.33%	0.58%	0.40%	0.63%	0.20%	0.26%	0.12%	0.52%	0.51%	0.06%	0.01%	0.20%	0.13%	0.25%	0.06%	0.02%	0.00%	0.00%	0.00%	0.03%	0.04%	0.24%
й	0.03%	0.03%	0.07%	0.03%	0.07%	0.01%	0.00%	0.02%	0.02%	0.07%	0.00%	0.06%	0.03%	0.04%	0.10%	0.06%	0.08%	0.03%	0.17%	0.06%	0.02%	0.01%	0.01%	0.01%	0.03%	0.02%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.01%
к	0.89%	0.05%	0.07%	0.01%	0.03%	0.06%	0.00%	0.02%	0.01%	0.34%	0.00%	0.04%	0.05%	0.02%	0.19%	0.99%	0.04%	0.18%	0.11%	0.07%	0.18%	0.00%	0.01%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.01%
л	0.84%	0.03%	0.08%	0.04%	0.05%	0.60%	0.00%	0.05%	0.02%	0.75%	0.00%	0.11%	0.04%	0.01%	0.11%	0.82%	0.09%	0.02%	0.22%	0.04%	0.15%	0.00%	0.00%	0.00%	0.07%	0.00%	0.00%	0.00%	0.12%	0.43%	0.01%	0.12%	0.22%
м	0.31%	0.03%	0.07%	0.03%	0.04%	0.35%	0.00%	0.01%	0.02%	0.42%	0.00%	0.06%	0.03%	0.03%	0.20%	0.46%	0.09%	0.03%	0.10%	0.04%	0.32%	0.00%	0.01%	0.00%	0.05%	0.01%	0.00%	0.00%	0.10%	0.01%	0.01%	0.00%	0.06%
н	1.27%	0.03%	0.06%	0.02%	0.11%	1.11%	0.00%	0.01%	0.02%	0.95%	0.00%	0.05%	0.01%	0.01%	0.40%	1.12%	0.07%	0.02%	0.15%	0.07%	0.30%	0.00%	0.01%	0.06%	0.04%	0.00%	0.02%	0.00%	0.36%	0.13%	0.00%	0.02%	0.26%
о	0.02%	0.54%	1.16%	0.61%	0.65%	0.37%	0.00%	0.27%	0.18%	0.24%	0.39%	0.31%	0.78%	0.65%	1.14%	0.27%	0.38%	0.76%	1.00%	0.95%	0.07%	0.04%	0.07%	0.01%	0.30%	0.15%	0.02%	0.00%	0.00%	0.00%	0.05%	0.08%	0.11%
п	0.14%	0.00%	0.00%	0.00%	0.00%	0.27%	0.00%	0.00%	0.00%	0.09%	0.00%	0.01%	0.07%	0.00%	0.01%	1.01%	0.00%	0.74%	0.00%	0.01%	0.07%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.03%	0.06%	0.00%	0.00%	0.04%
р	0.90%	0.02%	0.04%	0.04%	0.04%	0.66%	0.00%	0.04%	0.01%	0.58%	0.00%	0.05%	0.01%	0.03%	0.09%	0.80%	0.02%	0.00%	0.04%	0.06%	0.32%	0.00%	0.01%	0.01%	0.01%	0.03%	0.00%	0.00%	0.17%	0.11%	0.00%	0.01%	0.12%
с	0.19%	0.02%	0.23%	0.01%	0.05%	0.45%	0.02%	0.02%	0.01%	0.18%	0.00%	0.60%	0.30%	0.12%	0.14%	0.33%	0.23%	0.04%	0.12%	1.31%	0.09%	0.00%	0.02%	0.00%	0.05%	0.01%	0.00%	0.00%	0.04%	0.34%	0.01%	0.01%	0.37%
т	0.62%	0.03%	0.36%	0.01%	0.04%	0.57%	0.00%	0.01%	0.01%	0.47%	0.00%	0.08%	0.04%	0.02%	0.17%	1.76%	0.05%	0.31%	0.16%	0.04%	0.17%	0.00%	0.01%	0.01%	0.05%	0.00%	0.00%	0.00%	0.17%	0.68%	0.01%	0.01%	0.07%
у	0.02%	0.09%	0.18%	0.15%	0.23%	0.04%	0.00%	0.18%	0.09%	0.07%	0.01%	0.15%	0.15%	0.15%	0.12%	0.05%	0.14%	0.08%	0.20%	0.18%	0.01%	0.00%	0.04%	0.00%	0.14%	0.08%	0.03%	0.00%	0.00%	0.00%	0.01%	0.11%	0.02%
ф	0.01%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.06%	0.00%	0.00%	0.01%	0.00%	0.00%	0.01%	0.00%	0.05%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
х	0.10%	0.01%	0.04%	0.01%	0.02%	0.01%	0.00%	0.00%	0.01%	0.05%	0.00%	0.02%	0.03%	0.02%	0.04%	0.28%	0.04%	0.02%	0.05%	0.01%	0.03%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ц	0.06%	0.00%	0.01%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.03%	0.00%	0.01%	0.00%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%
ч	0.25%	0.00%	0.01%	0.00%	0.00%	0.38%	0.00%	0.00%	0.00%	0.17%	0.00%	0.03%	0.00%	0.00%	0.06%	0.01%	0.01%	0.01%	0.01%	0.49%	0.09%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.00%
ш	0.14%	0.00%	0.01%	0.00%	0.00%	0.27%	0.00%	0.00%	0.00%	0.21%	0.00%	0.04%	0.06%	0.00%	0.04%	0.03%	0.00%	0.00%	0.00%	0.01%	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%
щ	0.04%	0.00%	0.00%	0.00%	0.00%	0.15%	0.00%	0.00%	0.00%	0.08%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ъ	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%
ы	0.01%	0.07%	0.17%	0.02%	0.04%	0.14%	0.00%	0.01%	0.02%	0.05%	0.15%	0.04%	0.30%	0.18%	0.09%	0.03%	0.06%	0.05%	0.14%	0.10%	0.01%	0.00%	0.11%	0.00%	0.03%	0.05%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%
ь	0.03%	0.04%	0.12%	0.03%	0.05%	0.16%	0.00%	0.01%	0.05%	0.12%	0.00%	0.21%	0.02%	0.06%	0.24%	0.09%	0.10%	0.03%	0.22%	0.06%	0.02%	0.00%	0.01%	0.01%	0.06%	0.06%	0.00%	0.00%	0.00%	0.00%	0.03%	0.05%	0.06%
э	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ю	0.01%	0.06%	0.03%	0.01%	0.06%	0.01%	0.00%	0.01%	0.01%	0.03%	0.00%	0.02%	0.01%	0.01%	0.03%	0.03%	0.03%	0.02%	0.05%	0.06%	0.01%	0.00%	0.00%	0.00%	0.02%	0.01%	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
я	0.03%	0.05%	0.16%	0.06%	0.12%	0.05%	0.00%	0.06%	0.11%	0.11%	0.01%	0.10%	0.12%	0.08%	0.24%	0.09%	0.11%	0.03%	0.22%	0.23%	0.03%	0.00%	0.03%	0.01%	0.05%	0.01%	0.02%	0.00%	0.00%	0.00%	0.02%	0.01%	0.03%

Табл. 4: Таблица біграм для тексту без пробілів (біграми 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	0.00%	0.06%	0.32%	0.07%	0.21%	0.10%	0.00%	0.12%	0.40%	0.01%	0.05%	0.40%	0.90%	0.25%	0.44%	0.00%	0.08%	0.26%	0.35%	0.43%	0.01%	0.04%	0.09%	0.01%	0.07%	0.09%	0.02%	0.00%	0.00%	0.00%	0.00%	0.08%	0.24%	1.72%
б	0.09%	0.00%	0.01%	0.00%	0.00%	0.20%	0.00%	0.00%	0.00%	0.07%	0.00%	0.02%	0.07%	0.00%	0.03%	0.20%	0.00%	0.11%	0.01%	0.00%	0.10%	0.00%	0.01%	0.00%	0.00%	0.02%	0.01%	0.40%	0.00%	0.00%	0.00%	0.05%	0.03%	
в	0.54%	0.00%	0.00%	0.00%	0.03%	0.43%	0.00%	0.00%	0.05%	0.38%	0.00%	0.01%	0.09%	0.01%	0.12%	0.72%	0.02%	0.08%	0.32%	0.02%	0.07%	0.00%	0.00%	0.00%	0.01%	0.12%	0.00%	0.00%	0.24%	0.02%	0.00%	0.02%	0.58%	
г	0.09%	0.00%	0.00%	0.00%	0.11%	0.05%	0.00%	0.00%	0.00%	0.07%	0.00%	0.01%	0.15%	0.00%	0.02%	0.85%	0.00%	0.09%	0.00%	0.00%	0.06%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.08%	
д	0.42%	0.00%	0.09%	0.00%	0.00%	0.44%	0.00%	0.00%	0.00%	0.23%	0.00%	0.02%	0.08%	0.01%	0.16%	0.36%	0.01%	0.16%	0.04%	0.02%	0.16%	0.00%	0.01%	0.02%	0.00%	0.01%	0.00%	0.01%	0.06%	0.06%	0.00%	0.04%	0.10%	
е	0.00%	0.12%	0.19%	0.35%	0.25%	0.18%	0.00%	0.09%	0.12%	0.01%	0.24%	0.13%	0.58%	0.39%	0.73%	0.04%	0.11%	0.59%	0.37%	0.39%	0.01%	0.00%	0.09%	0.02%	0.08%	0.08%	0.06%	0.00%	0.00%	0.00%	0.03%	0.03%	1.79%	
ё	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	
ж	0.14%	0.00%	0.00%	0.00%	0.08%	0.38%	0.00%	0.00%	0.00%	0.14%	0.00%	0.01%	0.00%	0.00%	0.11%	0.00%	0.00%	0.00%	0.01%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	
з	0.53%	0.02%	0.08%	0.04%	0.07%	0.02%	0.00%	0.01%	0.00%	0.03%	0.00%	0.01%	0.02%	0.03%	0.17%	0.06%	0.00%	0.02%	0.01%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.05%	0.04%	0.00%	0.00%	0.05%	0.12%	
и	0.01%	0.04%	0.25%	0.04%	0.15%	0.26%	0.00%	0.04%	0.18%	0.07%	0.13%	0.17%	0.44%	0.28%	0.35%	0.01%	0.02%	0.05%	0.25%	0.33%	0.00%	0.00%	0.16%	0.11%	0.15%	0.04%	0.01%	0.00%	0.00%	0.00%	0.03%	0.18%	1.88%	
й	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.03%	0.00%	0.00%	0.00%	0.06%	0.02%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.72%	
к	0.73%	0.00%	0.03%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.26%	0.00%	0.00%	0.03%	0.00%	0.10%	0.79%	0.00%	0.15%	0.04%	0.04%	0.14%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.46%	
л	0.69%	0.00%	0.00%	0.01%	0.02%	0.46%	0.00%	0.04%	0.00%	0.58%	0.00%	0.04%	0.02%	0.00%	0.03%	0.60%	0.01%	0.00%	0.12%	0.00%	0.11%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.10%	0.36%	0.00%	0.10%	0.18%	0.73%	
м	0.25%	0.00%	0.00%	0.00%	0.00%	0.28%	0.00%	0.00%	0.00%	0.29%	0.00%	0.01%	0.01%	0.00%	0.11%	0.34%	0.01%	0.01%	0.01%	0.00%	0.25%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.08%	0.00%	0.00%	0.00%	0.04%	0.72%	
н	1.04%	0.00%	0.00%	0.01%	0.07%	0.92%	0.00%	0.00%	0.00%	0.76%	0.00%	0.02%	0.00%	0.00%	0.27%	0.91%	0.00%	0.01%	0.07%	0.04%	0.23%	0.00%	0.00%	0.05%	0.02%	0.00%	0.01%	0.00%	0.30%	0.10%	0.00%	0.01%	0.22%	0.47%
о	0.00%	0.33%	0.74%	0.46%	0.44%	0.24%	0.00%	0.18%	0.10%	0.07%	0.32%	0.16%	0.61%	0.47%	0.72%	0.02%	0.11%	0.58%	0.64%	0.64%	0.01%	0.03%	0.04%	0.01%	0.14%	0.12%	0.01%	0.00%	0.00%	0.00%	0.06%	0.06%	2.31%	
п	0.12%	0.00%	0.00%	0.00%	0.00%	0.22%	0.00%	0.00%	0.00%	0.07%	0.00%	0.01%	0.06%	0.00%	0.01%	0.84%	0.00%	0.62%	0.00%	0.01%	0.06%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.03%	0.05%	0.00%	0.00%	0.03%	0.00%
р	0.75%	0.02%	0.03%	0.03%	0.03%	0.55%	0.00%	0.03%	0.00%	0.47%	0.00%	0.04%	0.01%	0.02%	0.06%	0.66%	0.01%	0.00%	0.02%	0.05%	0.26%	0.00%	0.01%	0.00%	0.01%	0.03%	0.00%	0.00%	0.14%	0.09%	0.00%	0.01%	0.10%	0.10%
с	0.16%	0.00%	0.17%	0.00%	0.03%	0.37%	0.01%	0.00%	0.00%	0.14%	0.00%	0.48%	0.24%	0.09%	0.08%	0.26%	0.17%	0.02%	0.07%	1.06%	0.07%	0.00%	0.02%	0.00%	0.03%	0.01%	0.00%	0.03%	0.29%	0.00%	0.01%	0.30%	0.28%	
т	0.51%	0.00%	0.26%	0.00%	0.01%	0.46%	0.00%	0.00%	0.00%	0.37%	0.00%	0.04%	0.02%	0.00%	0.10%	1.44%	0.01%	0.25%	0.09%	0.01%	0.13%	0.00%	0.00%	0.01%	0.02%	0.00%	0.00%	0.14%	0.57%	0.00%	0.01%	0.06%	0.44%	
у	0.00%	0.05%	0.10%	0.11%	0.16%	0.02%	0.00%	0.15%	0.06%	0.00%	0.01%	0.08%	0.12%	0.10%	0.02%	0.00%	0.06%	0.05%	0.12%	0.12%	0.00%	0.00%	0.03%	0.00%	0.08%	0.07%	0.02%	0.00%	0.00%	0.00%	0.09%	0.01%	0.62%	
ф	0.01%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.04%	0.00%	0.00%	0.01%	0.00%	0.00%	0.01%	0.00%	0.04%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	
х	0.08%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.01%	0.00%	0.01%	0.22%	0.00%	0.01%	0.01%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.31%	
ц	0.05%	0.00%	0.00%	0.00%	0.00%	0.08%	0.00%	0.00%	0.00%	0.02%	0.00%	0.01%	0.00%	0.00%	0.00%	0.03%	0.00%	0.00%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.02%	
ч	0.21%	0.00%	0.00%	0.00%	0.00%	0.32%	0.00%	0.00%	0.00%	0.14%	0.00%	0.03%	0.00%	0.00%	0.05%	0.01%	0.00%	0.00%	0.00%	0.40%	0.07%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.02%	0.00%	0.00%	0.00%	0.05%	
ш	0.12%	0.00%	0.00%	0.00%	0.00%	0.22%	0.00%	0.00%	0.00%	0.17%	0.00%	0.03%	0.05%	0.00%	0.03%	0.03%	0.00%	0.00%	0.00%	0.01%	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.01%	
щ	0.03%	0.00%	0.00%	0.00%	0.00%	0.12%	0.00%	0.00%	0.00%	0.07%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
ъ	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.00%	
ы	0.00%	0.04%	0.11%	0.01%	0.01%	0.10%	0.00%	0.00%	0.01%	0.00%	0.13%	0.02%	0.25%	0.13%	0.03%	0.00%	0.01%	0.03%	0.08%	0.07%	0.00%	0.00%	0.09%	0.00%	0.02%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.42%	
ь	0.00%	0.01%	0.00%	0.00%	0.00%	0.09%	0.00%	0.00%	0.01%	0.02%	0.00%	0.11%	0.00%	0.02%	0.11%	0.00%	0.00%	0.00%	0.08%	0.01%	0.00%	0.00%	0.00%	0.01%	0.01%	0.05%	0.00%	0.00%	0.00%	0.00%	0.04%	0.04%	1.00%	
э	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.25%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
ю	0.00%	0.05%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.04%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.31%		
я	0.00%	0.00%	0.03%	0.02%	0.06%	0.01%	0.00%	0.04%	0.06%	0.00%	0.01%	0.01%	0.08%	0.03%	0.06%	0.00%	0.00%	0.00%	0.08%	0.14%	0.00%	0.00%	0.02%	0.00%	0.01%	0.00%	0.02%	0.00%	0.00%	0.00%	0.01%	0.01%	1.08%	
—	0.26%	0.69%	1.47%	0.41%	0.74%	0.46%	0.00%	0.22%	0.39%	1.12%	0.00%	0.96%	0.31%	0.54%	1.60%	1.27%	1.49%	0.40%	1.58%	0.80%	0.39%	0.06%	0.12%	0.03%	0.58%	0.08%	0.01%	0.00%	0.00%	0.27%	0.00%	0.17%	0.25%	

Табл. 5: Таблица частот біграм для тексту з пробілами (біграми 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	0.04%	0.15%	0.57%	0.13%	0.34%	0.19%	0.00%	0.16%	0.52%	0.16%	0.06%	0.62%	1.14%	0.37%	0.75%	0.17%	0.28%	0.35%	0.61%	0.62%	0.06%	0.05%	0.12%	0.01%	0.15%	0.12%	0.03%	0.00%	0.00%	0.00%	0.03%	0.10%	0.31%
б	0.11%	0.00%	0.01%	0.00%	0.00%	0.25%	0.00%	0.00%	0.00%	0.09%	0.00%	0.03%	0.09%	0.01%	0.03%	0.26%	0.00%	0.13%	0.01%	0.00%	0.12%	0.00%	0.01%	0.00%	0.00%	0.00%	0.02%	0.01%	0.48%	0.00%	0.01%	0.00%	0.05%
в	0.65%	0.02%	0.04%	0.04%	0.07%	0.53%	0.00%	0.01%	0.07%	0.48%	0.00%	0.08%	0.13%	0.04%	0.19%	0.92%	0.10%	0.12%	0.46%	0.08%	0.10%	0.00%	0.01%	0.00%	0.02%	0.15%	0.00%	0.00%	0.29%	0.02%	0.03%	0.00%	0.03%
г	0.11%	0.00%	0.01%	0.00%	0.13%	0.06%	0.00%	0.00%	0.00%	0.09%	0.00%	0.02%	0.17%	0.00%	0.03%	1.03%	0.01%	0.11%	0.02%	0.00%	0.08%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
д	0.51%	0.01%	0.12%	0.01%	0.01%	0.53%	0.00%	0.00%	0.00%	0.29%	0.00%	0.03%	0.10%	0.01%	0.21%	0.44%	0.02%	0.20%	0.06%	0.03%	0.20%	0.00%	0.01%	0.02%	0.01%	0.01%	0.00%	0.02%	0.07%	0.07%	0.00%	0.00%	0.05%
е	0.02%	0.25%	0.42%	0.47%	0.40%	0.26%	0.00%	0.14%	0.21%	0.17%	0.29%	0.26%	0.75%	0.59%	1.04%	0.20%	0.34%	0.76%	0.66%	0.57%	0.07%	0.01%	0.13%	0.02%	0.17%	0.10%	0.07%	0.00%	0.00%	0.00%	0.03%	0.04%	0.06%
ё	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ж	0.16%	0.01%	0.00%	0.00%	0.09%	0.46%	0.00%	0.00%	0.00%	0.16%	0.00%	0.01%	0.00%	0.00%	0.14%	0.01%	0.00%	0.00%	0.01%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
з	0.63%	0.02%	0.11%	0.06%	0.09%	0.03%	0.00%	0.02%	0.01%	0.04%	0.00%	0.03%	0.02%	0.04%	0.22%	0.07%	0.02%	0.03%	0.03%	0.01%	0.03%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.06%	0.05%	0.00%	0.00%	0.06%
и	0.03%	0.14%	0.54%	0.11%	0.28%	0.37%	0.00%	0.06%	0.27%	0.23%	0.15%	0.33%	0.57%	0.40%	0.63%	0.20%	0.25%	0.12%	0.52%	0.51%	0.06%	0.01%	0.20%	0.13%	0.25%	0.05%	0.02%	0.00%	0.00%	0.00%	0.03%	0.04%	0.24%
й	0.03%	0.03%	0.07%	0.03%	0.08%	0.01%	0.00%	0.01%	0.02%	0.07%	0.00%	0.06%	0.03%	0.04%	0.10%	0.05%	0.08%	0.03%	0.17%	0.06%	0.02%	0.01%	0.01%	0.01%	0.04%	0.02%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.01%
к	0.88%	0.04%	0.07%	0.01%	0.03%	0.06%	0.00%	0.02%	0.01%	0.34%	0.00%	0.04%	0.05%	0.02%	0.19%	0.99%	0.04%	0.18%	0.11%	0.08%	0.18%	0.00%	0.01%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.01%
л	0.82%	0.03%	0.08%	0.04%	0.05%	0.60%	0.00%	0.05%	0.02%	0.75%	0.00%	0.11%	0.04%	0.01%	0.11%	0.82%	0.08%	0.02%	0.22%	0.04%	0.15%	0.00%	0.00%	0.00%	0.07%	0.00%	0.00%	0.00%	0.12%	0.43%	0.02%	0.12%	0.21%
м	0.31%	0.03%	0.07%	0.03%	0.04%	0.35%	0.00%	0.01%	0.02%	0.42%	0.00%	0.06%	0.03%	0.03%	0.20%	0.46%	0.09%	0.03%	0.10%	0.04%	0.32%	0.00%	0.01%	0.00%	0.06%	0.01%	0.00%	0.00%	0.11%	0.01%	0.01%	0.00%	0.06%
н	1.28%	0.03%	0.06%	0.02%	0.11%	1.11%	0.00%	0.01%	0.02%	0.96%	0.00%	0.05%	0.01%	0.01%	0.39%	1.12%	0.07%	0.02%	0.14%	0.07%	0.29%	0.00%	0.01%	0.06%	0.04%	0.00%	0.01%	0.00%	0.36%	0.13%	0.00%	0.02%	0.26%
о	0.02%	0.54%	1.16%	0.62%	0.65%	0.37%	0.00%	0.27%	0.17%	0.24%	0.39%	0.31%	0.78%	0.65%	1.14%	0.27%	0.37%	0.76%	1.00%	0.96%	0.07%	0.04%	0.07%	0.01%	0.30%	0.15%	0.02%	0.00%	0.00%	0.00%	0.05%	0.07%	0.11%
п	0.14%	0.00%	0.00%	0.00%	0.00%	0.27%	0.00%	0.00%	0.00%	0.09%	0.00%	0.01%	0.08%	0.00%	0.01%	1.01%	0.00%	0.74%	0.00%	0.01%	0.07%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.03%	0.06%	0.00%	0.00%	0.04%
р	0.91%	0.02%	0.04%	0.04%	0.04%	0.67%	0.00%	0.04%	0.01%	0.57%	0.00%	0.05%	0.01%	0.03%	0.09%	0.81%	0.02%	0.00%	0.04%	0.06%	0.32%	0.00%	0.01%	0.01%	0.01%	0.03%	0.00%	0.00%	0.17%	0.10%	0.00%	0.02%	0.12%
с	0.19%	0.02%	0.24%	0.01%	0.05%	0.45%	0.02%	0.02%	0.01%	0.18%	0.00%	0.60%	0.31%	0.12%	0.14%	0.33%	0.22%	0.04%	0.12%	1.31%	0.09%	0.00%	0.02%	0.01%	0.05%	0.01%	0.00%	0.00%	0.04%	0.34%	0.01%	0.01%	0.37%
т	0.61%	0.03%	0.35%	0.01%	0.04%	0.56%	0.00%	0.01%	0.01%	0.48%	0.00%	0.08%	0.04%	0.02%	0.17%	1.76%	0.05%	0.31%	0.15%	0.04%	0.17%	0.00%	0.01%	0.01%	0.05%	0.00%	0.00%	0.00%	0.17%	0.68%	0.01%	0.01%	0.07%
у	0.02%	0.09%	0.18%	0.15%	0.23%	0.04%	0.00%	0.19%	0.09%	0.07%	0.01%	0.14%	0.15%	0.15%	0.12%	0.05%	0.13%	0.07%	0.20%	0.18%	0.01%	0.00%	0.04%	0.00%	0.14%	0.09%	0.03%	0.00%	0.00%	0.00%	0.01%	0.12%	0.02%
ф	0.01%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.00%	0.06%	0.00%	0.00%	0.01%	0.00%	0.00%	0.01%	0.00%	0.05%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
х	0.10%	0.02%	0.04%	0.01%	0.02%	0.01%	0.00%	0.00%	0.01%	0.05%	0.00%	0.02%	0.03%	0.02%	0.04%	0.28%	0.04%	0.02%	0.05%	0.01%	0.02%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ц	0.06%	0.00%	0.01%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.03%	0.00%	0.01%	0.00%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%
ч	0.25%	0.00%	0.01%	0.00%	0.00%	0.38%	0.00%	0.00%	0.00%	0.16%	0.00%	0.03%	0.00%	0.00%	0.06%	0.01%	0.01%	0.00%	0.01%	0.48%	0.09%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.00%
ш	0.14%	0.00%	0.01%	0.00%	0.00%	0.27%	0.00%	0.00%	0.00%	0.21%	0.00%	0.04%	0.06%	0.00%	0.04%	0.03%	0.00%	0.00%	0.00%	0.01%	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%
щ	0.04%	0.00%	0.00%	0.00%	0.00%	0.15%	0.00%	0.00%	0.00%	0.08%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ъ	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	
ы	0.01%	0.07%	0.17%	0.02%	0.04%	0.14%	0.00%	0.01%	0.02%	0.05%	0.15%	0.04%	0.30%	0.18%	0.10%	0.03%	0.06%	0.05%	0.14%	0.10%	0.01%	0.00%	0.11%	0.00%	0.03%	0.05%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%
ь	0.03%	0.05%	0.12%	0.03%	0.05%	0.16%	0.00%	0.01%	0.05%	0.12%	0.00%	0.21%	0.02%	0.06%	0.24%	0.09%	0.10%	0.03%	0.22%	0.06%	0.02%	0.00%	0.01%	0.01%	0.06%	0.06%	0.00%	0.00%	0.00%	0.00%	0.03%	0.05%	0.06%
э	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.31%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ю	0.01%	0.06%	0.03%	0.01%	0.06%	0.01%	0.00%	0.01%	0.01%	0.03%	0.00%	0.02%	0.01%	0.01%	0.03%	0.03%	0.03%	0.02%	0.05%	0.06%	0.01%	0.00%	0.00%	0.00%	0.02%	0.01%	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
я	0.03%	0.05%	0.16%	0.05%	0.12%	0.05%	0.00%	0.07%	0.11%	0.11%	0.01%	0.10%	0.12%	0.07%	0.24%	0.09%	0.12%	0.03%	0.22%	0.23%	0.02%	0.00%	0.03%	0.01%	0.05%	0.01%	0.03%	0.00%	0.00%	0.00%	0.02%	0.01%	0.03%

Табл. 6: Таблица біграм для тексту без пробілів (біграми з кроком 2)

### 3.2 Знаходження ентропії $H_1$ і $H_2$

Ентропії різні для різних текстів:

- Для тексту з пробілами, маємо:
  1.  $H_1 = 4.3622017$
  2.  $H_2 = 3.9618611$
  3.  $H'_2 = 3.9615$  - для біграм з кроком 2
- Для тексту без пробілів:
  1.  $H_1 = 4.4544373$
  2.  $H_2 = 4.135257$
  3.  $H'_2 = 4.1351895$  - для біграм з кроком 2

### 3.3 Обчислення значень $H^{(10)}$ , $H^{(20)}$ , $H^{(30)}$

Ось результати експериментів в CoolPinkProgram:

Лабораторная работа №1

Произвольная часть текста:  
ли\_что\_ем

Использованные буквы:

Порядок n-граммы:  
5 символов

Введенный символ:

Символ по счету:

Номер эксперимента: 52

Неравенство для энтропии:  
 $3.09241379436319 < H < 3.62492760888188$

Двоичная таблица угаданных символов:

Вероятности:

q[1]	= 0.2745098
q[2]	= 0.0980392
q[3]	= 0.0392156
q[4]	= 0.0784313
q[5]	= 0.0588235
q[6]	= 0.0196078
q[7]	= 0
q[8]	= 0.0784313
q[9]	= 0.0196078
q[10]	= 0.019607
q[11]	= 0
q[12]	= 0.039215
q[13]	= 0.058823
q[14]	= 0.019607
q[15]	= 0
q[16]	= 0.019607
q[17]	= 0
q[18]	= 0.039215
q[19]	= 0
q[20]	= 0.078431
q[21]	= 0
q[22]	= 0.019607
q[23]	= 0
q[24]	= 0
q[25]	= 0.019607
q[26]	= 0.019607
q[27]	= 0
q[28]	= 0
q[29]	= 0
q[30]	= 0
q[31]	= 0
q[32]	= 0

Строка состояния:

Рис. 1: Значения для  $H^{(10)}$



З цих результатів отримуємо:

$$\begin{aligned} 3.0924 &\leq H^{(10)} \leq 3.6249 \\ 2.3964 &\leq H^{(20)} \leq 3.1001 \\ 2.7774 &\leq H^{(30)} \leq 3.4643 \end{aligned}$$

Порівнявши ці значення з значеннями  $H_1$  і  $H_2$  можна помітити, що ці значення менше.

### 3.4 Надлишковість мови

З отриманих значень від CoolPinkProgram можемо порахувати надлишковість російської мови за формулою  $R = 1 - \frac{H_{\text{inf}}}{H_0}$ , де  $H_0 = \log_2 32 = 5$  а замість  $H_{\text{inf}} = H^{(n)}$ :

$$\begin{aligned} \text{Для } H^{(10)}: 0.3815 &\geq R \geq 0.2750 \\ \text{Для } H^{(20)} 0.5207 &\geq R \geq 0.3799 \\ \text{Для } H^{(30)} 0.4445 &\geq R \geq 0.3071 \end{aligned}$$

Також порахуємо надлишковість для наших значень ентропії над алфавітом з пробілами:

$$\begin{aligned} \text{Для } H_1: R &= 0.1425 \\ \text{Для } H_2: R &= 0.2212 \end{aligned}$$

## 4 Висновок

Я навчився визначати ентропію на символ джерела, ентропію n-грами, та дізнався що таке надлишковість джерела відкритого тексту. Також навчився працювати з CoolPinkProgram, і отримувати гарну практику, завищенні оцінки ентропії та access violation.