

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         temp = temp.TrimStart();
15         return temp;
16     }
17     public static void ReformatFile(string inputFile, string outputFile, bool
18     withSpaces = true)
19     {
20         var reader = new StreamReader(inputFile, Encoding.UTF8);
21         using StreamWriter sw = new StreamWriter(outputFile);
22         while (!reader.EndOfStream)
23         {
24             string? line = reader.ReadLine();
25             if (line == null) break;
26             sw.Write(ReformatString(line, withSpaces));
27         }
28     }
29 }
```

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 енкодера ніяк не допомогала, але рішення виявилось доволі простим: треба було змінити енкодинг 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 }
```

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
—	595800	0.1645
о	349938	0.0966
е	256299	0.0708
а	247726	0.0684
и	204783	0.0565
н	201791	0.0557
т	179227	0.0495
с	160514	0.0443
л	152184	0.0420
в	141509	0.0391
р	128159	0.0354
к	102810	0.0284
д	91629	0.0253
м	87872	0.0243
у	82543	0.0228
п	77502	0.0214
я	66064	0.0182
ь	59161	0.0163
ы	57627	0.0159
г	57603	0.0159
б	52430	0.0145
з	50903	0.0141
ч	47066	0.0130
ж	33341	0.0092
й	32456	0.0090
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%	0.26%	0.45%	0.00%	0.08%	0.26%	0.35%	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%	1.73%
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%	0.59%
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%	0.86%	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%	0.73%	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%	0.19%	0.07%	0.13%	0.17%	0.45%	0.28%	0.35%	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%	1.88%
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%	0.00%	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	0.74%	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%	0.59%	0.00%	0.04%	0.02%	0.00%	0.03%	0.60%	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%	0.77%	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%	0.11%	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%	0.59%	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%	1.07%	0.07%	0.00%	0.02%	0.00%	0.04%	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%	0.36%	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%	0.45%
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%	0.63%
ф	0.01%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.05%	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%	0.14%	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%	0.00%	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%	0.43%	
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%	1.01%	
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%	0.70%	1.48%	0.41%	0.74%	0.46%	0.00%	0.22%	0.40%	1.12%	0.00%	0.95%	0.31%	0.54%	1.59%	1.28%	1.49%	0.40%	1.58%	0.81%	0.39%	0.07%	0.13%	0.03%	0.59%	0.08%	0.01%	0.00%	0.00%	0.27%	0.00%	0.17%	0.00%	

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%	0.20%	0.10%	0.00%	0.12%	0.40%	0.01%	0.05%	0.41%	0.90%	0.25%	0.45%	0.00%	0.08%	0.26%	0.34%	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%	0.21%	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%	0.04%	0.03%
в	0.54%	0.00%	0.00%	0.00%	0.03%	0.42%	0.00%	0.00%	0.05%	0.37%	0.00%	0.01%	0.09%	0.01%	0.12%	0.72%	0.02%	0.08%	0.31%	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%	0.60%
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%	0.86%	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%	0.17%	0.03%	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%	0.59%	0.39%	0.73%	0.04%	0.11%	0.59%	0.37%	0.39%	0.01%	0.00%	0.08%	0.02%	0.09%	0.08%	0.06%	0.00%	0.00%	0.00%	0.00%	0.03%	0.03%	1.77%
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	0.53%	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%	0.03%	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%	0.05%	0.15%	0.26%	0.00%	0.04%	0.18%	0.07%	0.13%	0.17%	0.45%	0.29%	0.35%	0.01%	0.02%	0.05%	0.25%	0.34%	0.00%	0.00%	0.15%	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%	0.04%	0.00%	0.11%	0.80%	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%	0.47%	0.00%	0.04%	0.00%	0.58%	0.00%	0.04%	0.02%	0.00%	0.03%	0.60%	0.01%	0.00%	0.11%	0.00%	0.12%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.00%	0.10%	0.36%	0.00%	0.10%	0.18%	0.71%
м	0.26%	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%	0.10%	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%	0.01%	0.00%	0.00%	0.04%	0.72%
н	1.05%	0.00%	0.00%	0.01%	0.07%	0.93%	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%	0.02%	0.21%	0.48%
о	0.00%	0.33%	0.75%	0.46%	0.45%	0.24%	0.00%	0.18%	0.10%	0.08%	0.33%	0.16%	0.61%	0.48%	0.73%	0.02%	0.12%	0.59%	0.63%	0.65%	0.01%	0.02%	0.04%	0.01%	0.14%	0.12%	0.01%	0.00%	0.00%	0.00%	0.06%	0.06%	2.31%	
п	0.11%	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%	0.83%	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%	0.02%	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%	0.47%	0.25%	0.09%	0.09%	0.26%	0.17%	0.02%	0.07%	1.07%	0.07%	0.00%	0.02%	0.00%	0.04%	0.01%	0.00%	0.04%	0.29%	0.00%	0.01%	0.30%	0.28%	
т	0.51%	0.00%	0.26%	0.00%	0.01%	0.45%	0.00%	0.00%	0.00%	0.36%	0.00%	0.04%	0.02%	0.00%	0.10%	1.45%	0.01%	0.25%	0.09%	0.01%	0.13%	0.00%	0.00%	0.01%	0.02%	0.00%	0.00%	0.14%	0.56%	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%	0.02%	0.00%	0.07%	0.05%	0.11%	0.13%	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%	0.05%	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%	0.21%	0.00%	0.01%	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%	0.32%	
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	0.20%	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%	0.08%	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%	0.02%	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%	0.01%	0.01%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.04%	0.04%	1.01%	
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%	0.01%	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%	0.00%	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%	0.09%	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%	0.02%	1.08%	
—	0.26%	0.70%	1.49%	0.40%	0.74%	0.47%	0.00%	0.22%	0.40%	1.13%	0.00%	0.95%	0.31%	0.55%	1.58%	1.28%	1.50%	0.40%	1.58%	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%	0.00%	

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. $H_1 = 4.3667064$
 2. $H_2 = 3.9591582$
 3. $H'_2 = 3.959069$ - 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:

The screenshot shows the CoolPinkProgram interface with the following data:

- Произвольная часть текста:** ли_что_ем
- Использованные буквы:** (empty)
- Порядок n-граммы:** 5 символов
- Введенный символ:** (empty)
- Символ по счету:** (empty)
- Номер эксперимента:** 52
- Неравенство для энтропии:** $3.09241379436319 < H < 3.62492760888188$
- Двоичная таблица угаданных символов:** (empty)
- 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

Рис. 1: Значення для $H^{(10)}$

Лабораторная работа №1

Произвольная часть текста:
еди_них_имеется_топ

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: 51

Поле ввода символов:

Продолжить Другой

Неравенство для энтропии:
2,39640298250587 < H < 3,10062938041322

Двоичная таблица угаданных символов:

Вероятности:

q[1] = 0.46
q[2] = 0.06
q[3] = 0
q[4] = 0.06
q[5] = 0.1
q[6] = 0.02
q[7] = 0.02
q[8] = 0.02
q[9] = 0
q[10] = 0
q[11] = 0.02
q[12] = 0.02
q[13] = 0
q[14] = 0
q[15] = 0.02
q[16] = 0.02
q[17] = 0.04
q[18] = 0
q[19] = 0.02
q[20] = 0.02
q[21] = 0
q[22] = 0.02
q[23] = 0
q[24] = 0
q[25] = 0.02
q[26] = 0
q[27] = 0.02
q[28] = 0
q[29] = 0
q[30] = 0.02
q[31] = 0.02
q[32] = 0

Строка состояния:

Рис. 2: Значения для $H^{(20)}$

Лабораторная работа №1

Произвольная часть текста:
сны_в_том_что_брать_каждо_по

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: 51

Поле ввода символов:

Продолжить Другой

Неравенство для энтропии:
2,77744371652263 < H < 3,46435048453636

Двоичная таблица угаданных символов:

Вероятности:

q[1] = 0.38
q[2] = 0.06
q[3] = 0.08
q[4] = 0.02
q[5] = 0.02
q[6] = 0.06
q[7] = 0.04
q[8] = 0
q[9] = 0.04
q[10] = 0
q[11] = 0.02
q[12] = 0.02
q[13] = 0
q[14] = 0
q[15] = 0
q[16] = 0.04
q[17] = 0.02
q[18] = 0.02
q[19] = 0
q[20] = 0.06
q[21] = 0
q[22] = 0
q[23] = 0.02
q[24] = 0.02
q[25] = 0
q[26] = 0
q[27] = 0.02
q[28] = 0.02
q[29] = 0
q[30] = 0
q[31] = 0.02
q[32] = 0.02

Строка состояния:

Рис. 3: Значения для $H^{(30)}$

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 $H_{\text{inf}} = H^{(30)}$, тоді:

$$H^{(30)}: 0.4445 \geq R \geq 0.3071$$

Тоді, виходить, що надлишковість мови: $R \approx 0.3758$

4 Висновок

Роблячи цю лабораторну, я зрозумів, як зчитувати кирилицю з текстових файлів, як вміщати великі таблиці в латеху, а також як важливо не поспішати і мати терпимість, для того щоб в CoolPinkProgram не вилізала помилка access violation.