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:

студенти ІІІ курсу ФТІ групи ФБ-96
Шидлюх Максим та Шафрай Ілля

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.

Завдання

1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H1 та H2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H1 та H2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H1 та H2 на тому ж тексті, в якому вилучено всі пробіли.

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

3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи

Завдання 1

В цій частині було визначено частоти простих і перехресних біграм. За цими показниками була визначена ентропія для біграм і літер. Частота біграм отсортована по їх буквенним значенням а для літер за зростанням частоти

Частоти літер

З пробілами	Без пробілів
0.1510455365672434," ", 0.06973781976435622,"е", 0.06729646534338181,"о", 0.06517354845557796,"а", 0.06283833987899373,"и", 0.051374588684852986,"в", 0.05052542192973145,"т", 0.041715316845345504,"н", 0.0404415667126632,"к", 0.038743233202420124,"л", 0.038212503980469166,"р", 0.02855323214096168,"с", 0.025368856809255918,"у", 0.024731981742914765,"ы", 0.024625835898524573,"п", 0.022078335633159963,"я", 0.019212397834624775,"м", *0.015815730814138628,"х",* *0.015815730814138628,"ш",* *0.015285001592187666,"ц",*	0.08214553638409602,"е", 0.0792698174543636,"о", 0.07676919229807452,"а", 0.07401850462615654,"и", 0.060515128782195546,"в", 0.05951487871967992,"т", 0.04913728432108027,"н", 0.047636909227306824,"к", 0.045636409102275566,"л", 0.0450112528132033,"р", 0.03363340835208802,"с", 0.029882470617654415,"у", 0.029132283070767692,"ы", 0.029007251812953237,"п", 0.022630657664416104,"м", 0.018629657414353587,"х", 0.018629657414353587,"ш", *0.01800450112528132,"ц",* 0.016129032258064516,"д", 0.015753938484621154,"з",

0.013692813926334785,"д", 0.013374376393164208,"з", 0.011357605349750556,"г", 0.011357605349750556,"ж", 0.010614584439019213,"й", 0.010296146905848636,"б", 0.010190001061458443,"ю", 0.008810105084385947,"ч", 0.007854792484874216,"ь", 0.005838021441460566 5,"щ", 0.004882708841948837,"ф", *0.0022290627321940345,"э",*	0.013378344586146536,"г", 0.013378344586146536,"ж", 0.012503125781445362,"й", 0.012128032008002,"б", 0.012003000750187547,"ю", 0.01037759439859965,"ч", 0.009252313078269568,"ь", 0.006876719179794949,"щ", 0.005751437859464866,"ф", *0.002625656414103526,"э",*

Частоти біграм(20 найчастіших)(перехресні)

З пробілами	Без пробілів
0.02635658914728682,"и ", 0.021705426356589147,"т ", 0.0212624584717608,"а ", 0.016611295681063124," в", 0.014285714285714285," п", 0.013953488372093023,"е ", 0.012846068660022148,"на", 0.012070874861572536," н", 0.011738648947951274,"о ", 0.011517165005537098,"ка", 0.011406423034330012,"в ", 0.010188261351052049," к", 0.009966777408637873,"ер", 0.00919158361018826,"ре", 0.009080841638981174," л", 0.008859357696566999,"он", 0.008637873754152824," о", 0.008527131782945736," и", 0.008416389811738648,"ки", 0.008084163898117386,"ло", *0.007862679955703212,"ст",*	0.015249112659392664,"на", 0.013671618246352045,"ка", 0.01327724464309189,"он", 0.011831208097804654,"ер", 0.011831208097804654,"ов", 0.01091100302353096,"ки", 0.01091100302353096,"ре", 0.010253713684764033,"ва", 0.009596424345997109,"ло", 0.009333508610490338,"ив", 0.009333508610490338,"ст", 0.009333508610490338,"во", 0.008544761403970027,"ры", 0.008544761403970027,"ев", 0.008413303536216643,"ни", *0.008413303536216643,"ле",* 0.008413303536216643,"ах", 0.008150387800709872,"ол", 0.008150387800709872,"ес", 0.008018929932956487,"ет", *0.007756014197449717 5,"ит",*

Частоти біграм(20 найчастіших)(прості)

З пробілами	Без пробілів
0.02823779193205945,"и ", 0.02335456475583864,"а ",	0.014615384615384615,"ка", 0.014358974358974359,"на",

0.02038216560509554," в",	0.013333333333333334,"он",
0.018683651804670912,"т ",	0.012564102564102564,"ки",
0.01592356687898089," н",	0.010769230769230769,"ре",
0.015286624203821656," п",	0.010769230769230769,"ер",
0.013163481953290871,"е ",	0.010769230769230769,"ов",
0.012314225053078557," о",	0.010512820512820513,"ва",
0.012314225053078557,"о ",	0.010256410256410256,"ло",
0.0118895966029724,"в ",	0.009743589743589744,"ив",
0.010615711252653927," к",	0.009743589743589744,"во",
0.010403397027600849,"ка",	0.009487179487179488,"ол",
0.009554140127388535," у",	0.009230769230769232,"ни",
0.009341825902335456," ж",	0.008717948717948718,"ст",
0.009129511677282378,"на",	0.008717948717948718,"ес",
0.009129511677282378,"ер",	0.008461538461538461,"ах",
0.00870488322717622,"ки",	0.008205128205128205,"ев",
0.00870488322717622,"ва",	0.008205128205128205,"ет",
0.00870488322717622,"ах",	0.007692307692307693,"не",
0.008492569002123142,"ры",	0.007692307692307693,"ти",

Ентропія для букв з пробілами: 4.444234149257989

Надлишковість: 0.1111531701484022

Ентропія для букв без пробілів : 4.428058475409482

Надлишковість: 0.09758361448440966

Ентропія для простих біграм без пробілів: 4.034882255155206

Надлишковість: 0.5310253963427229

Ентропія для простих біграм з пробілами: 3.859820556787146

Надлишковість: 0.5432819621103645

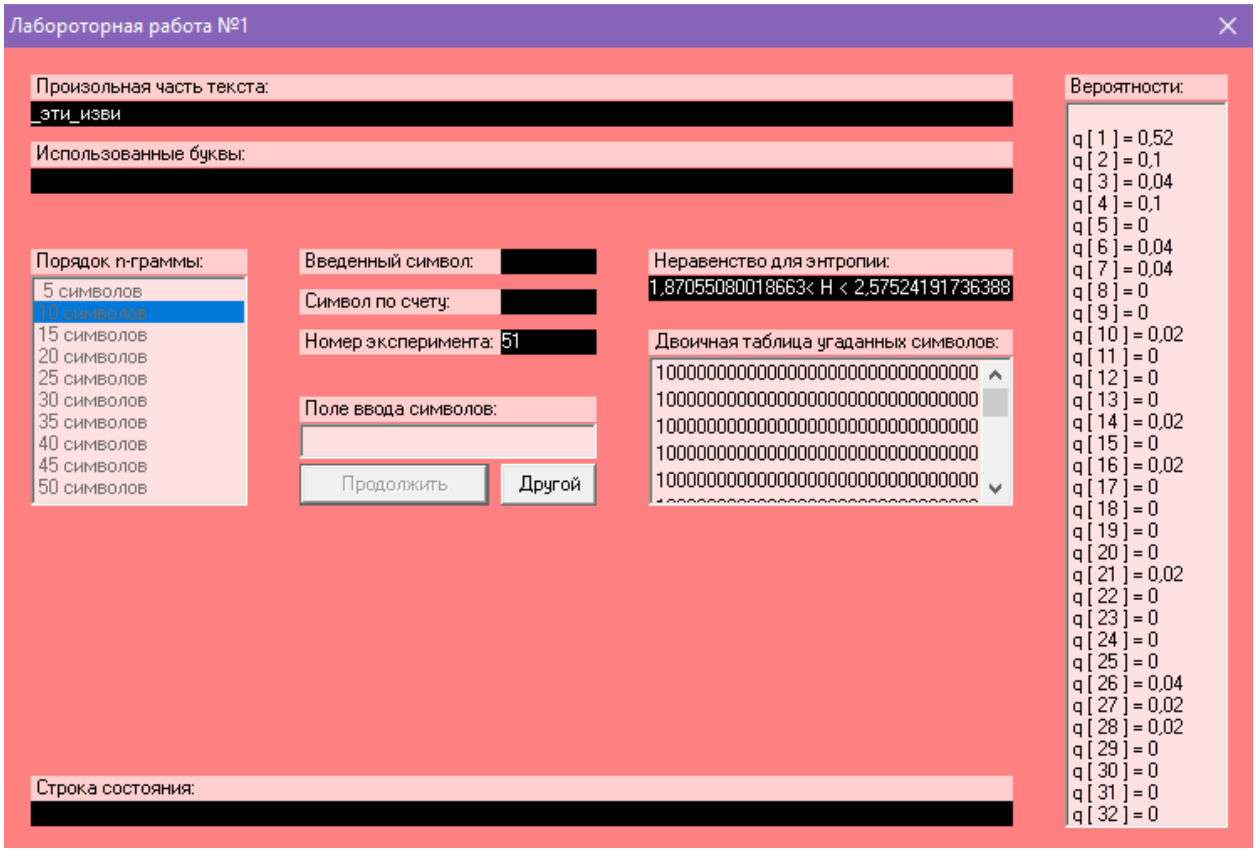
Ентропія для перехресних біграм без пробілів: 4.056088176091067

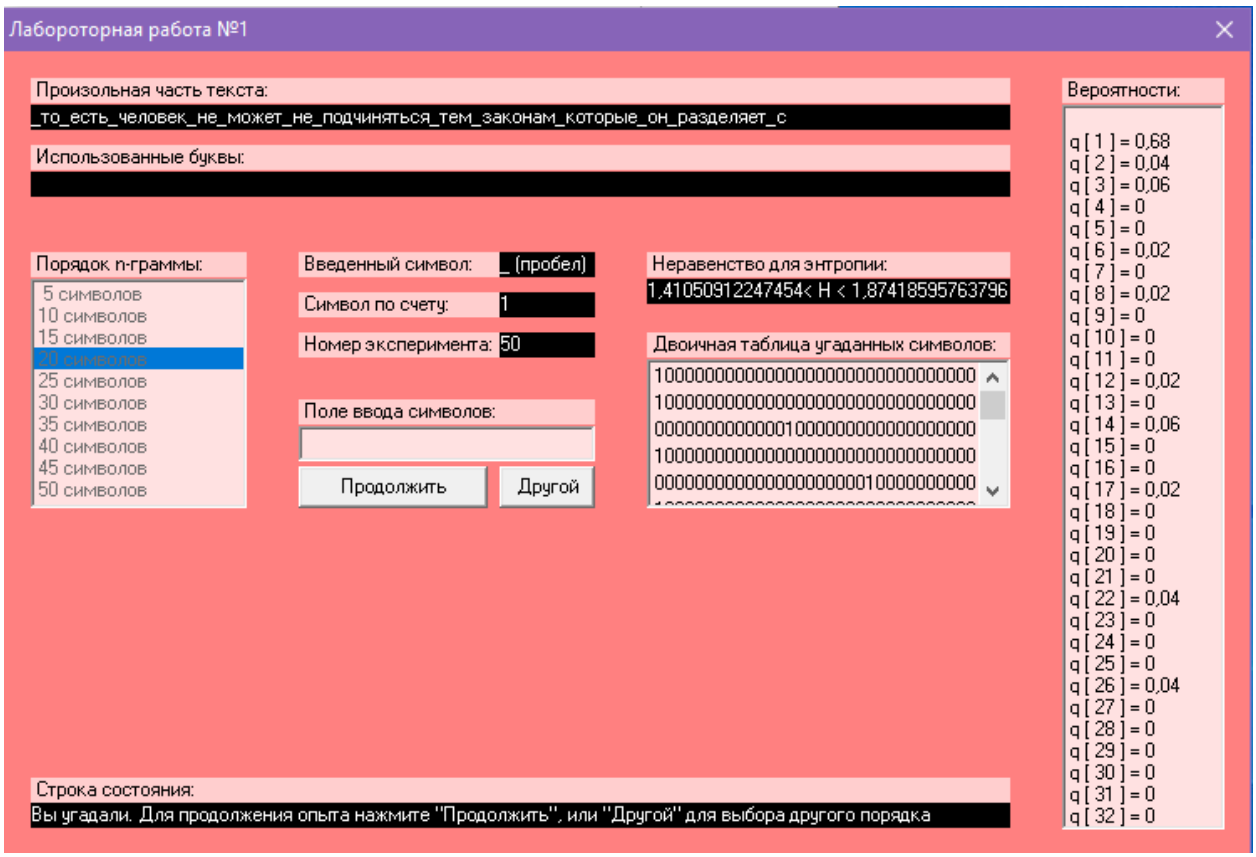
Надлишковість: 0.5387209156346204

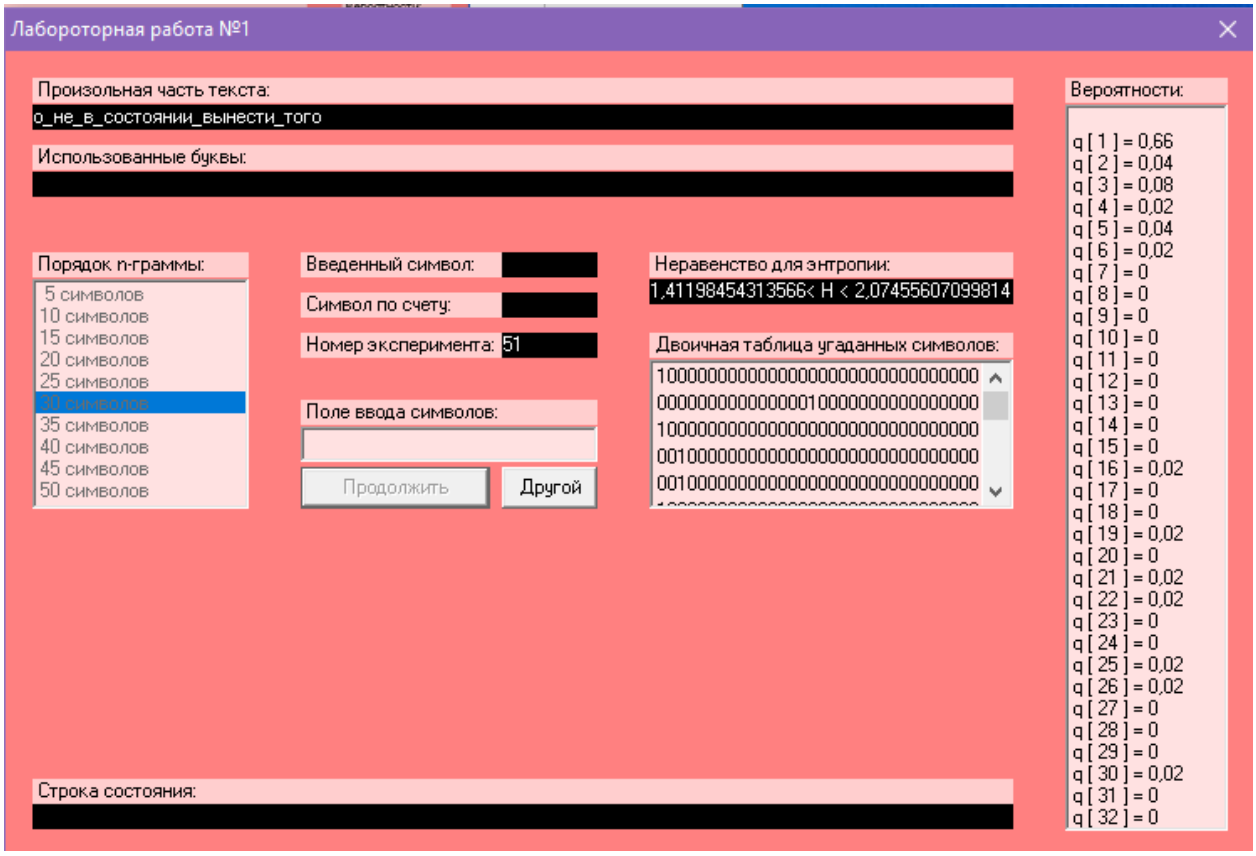
Ентропія для перехресних біграм з пробілами: 3.8983669234445384

Надлишковість: 0.5285606308955091

Завдання 2







Отримали значення

1.8705<H(10)<2.5752

1.4105<H(20)<1.8741

1.4119<H(30)<2.0745

Надлишковість :

```
import math
a = float(input(''))
print(1 - (a/math.log(32,2)))
```

0.6259>R(10)> 0.4849

0.7179>R(20)> 0.62518

0.7176>R(30)> 0.5851

Висновок:

Засвоїли поняття ентропії на символ джерела та його надлишковості, порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.