МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

Комп’ютерний практикум №1

з дисципліни Криптографія:

«Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Виконали:

Студенти групи ФБ-96

Шафрай Ілля

Шидлюх Максим

Київ 2021

Мета роботи

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

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

Частота літер

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

0.015815730814138628,"ш",	0.016129032258064516,"д",
0.015285001592187666,"ц",	0.015753938484621154,"з",
0.013692813926334785,"д",	0.013378344586146536,"г",
0.013374376393164208,"з",	0.013378344586146536,"ж",
0.011357605349750556,"г",	0.012503125781445362,"й",
0.011357605349750556,"ж",	0.012128032008002,"б",
0.010614584439019213,"й",	0.012003000750187547,"ю",
0.010296146905848636,"б",	0.01037759439859965,"ч",
0.010190001061458443,"ю",	0.009252313078269568,"ь",
0.008810105084385947,"ч",	0.006876719179794949,"щ",
0.007854792484874216,"ь",	0.005751437859464866,"ф",
0.0058380214414605665,"щ",	0.002625656414103526,"э",
0.004882708841948837,"ф",	
0.0022290627321940345,"э",	

Частота біграм (20 найчастіших)

З пробілами	Без пробілів
0.02635658914728682,"и ",	0.015249112659392664,"на",
0.021705426356589147,"т ",	0.013671618246352045,"ка",
0.0212624584717608,"а ",	0.01327724464309189,"он",
0.016611295681063124," в",	0.011831208097804654,"ер",
0.014285714285714285," п",	0.011831208097804654,"ов",
0.013953488372093023,"е ",	0.01091100302353096,"ки",
0.012846068660022148,"на",	0.01091100302353096,"ре",
0.012070874861572536," н",	0.010253713684764033,"ва",
0.011738648947951274,"о ",	0.009596424345997109,"ло",
0.011517165005537098,"ка",	0.009333508610490338,"ив",
0.011406423034330012,"в ",	0.009333508610490338,"ст",
0.010188261351052049," к",	0.009333508610490338,"во",
0.009966777408637873,"ер",	0.008544761403970027,"ры",
0.00919158361018826,"ре",	0.008544761403970027,"ев",
0.009080841638981174," л",	0.008413303536216643,"ни",
0.008859357696566999,"он",	0.008413303536216643,"ле",

0.008637873754152824," о",	0.008413303536216643,"ах",
0.00852713178294573 6," и",	0.008150387800709872,"ол",
0.008416389811738648,"ки",	0.008150387800709872,"ес",
0.008084163898117386,"ло",	0.008018929932956487,"ет",
0.007862679955703212,"ст",	0.0077560141974497175,"ит",

Ентропія для букв без пробілів: 4.353148523001946

Надлишковість: 0.8548950492332685

Ентропія для букв з пробілами: 4.419377566114722

Redundancy 0.8618944510589149

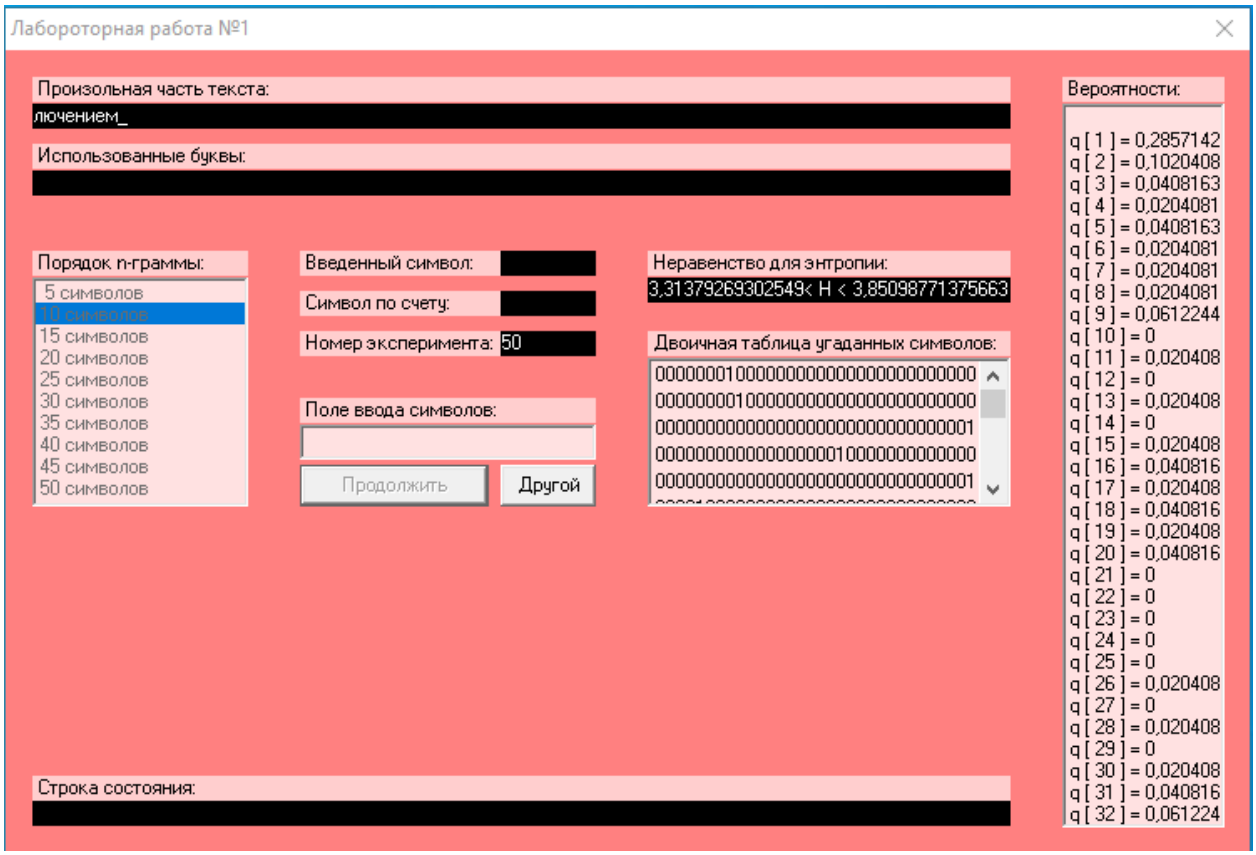
Ентропія для біграм без пробілів: 4.038541077351004

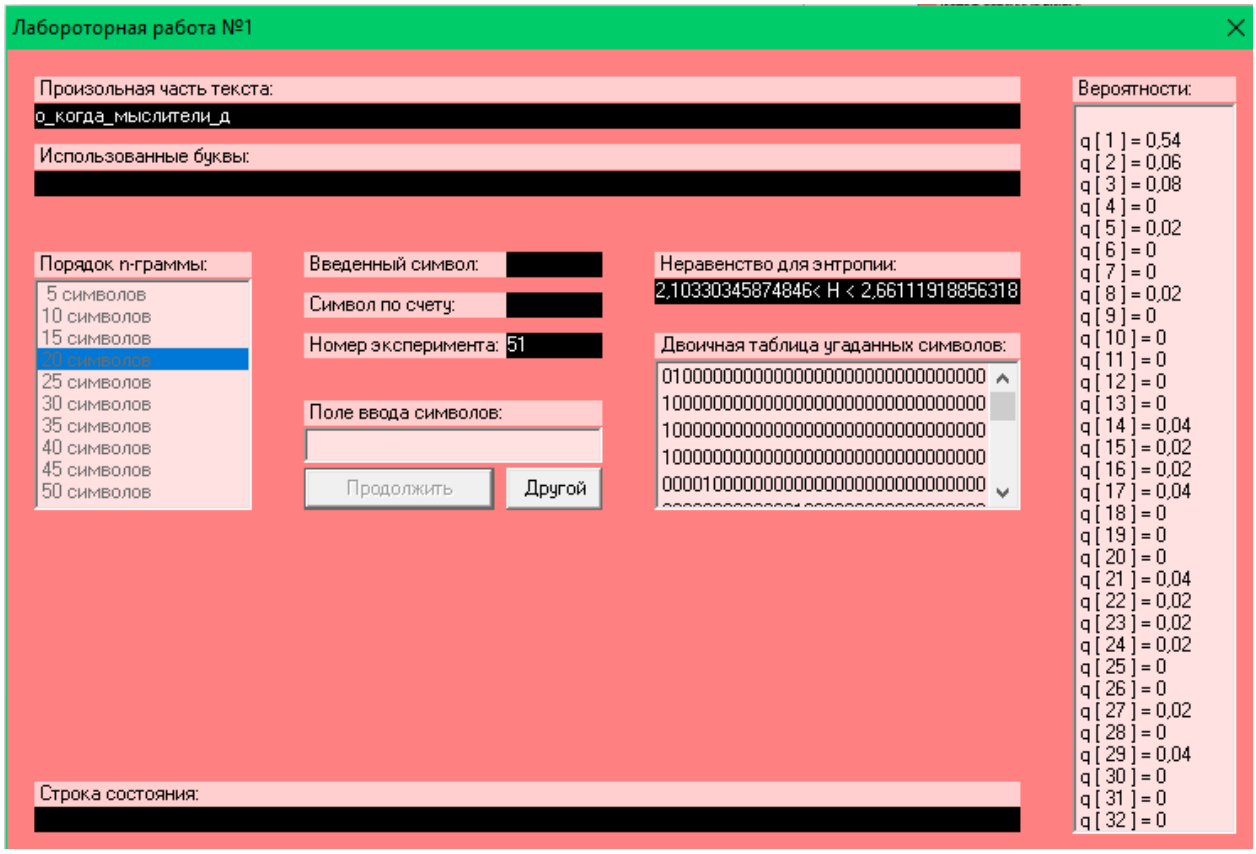
Надлишковість: 0.8653819640882998

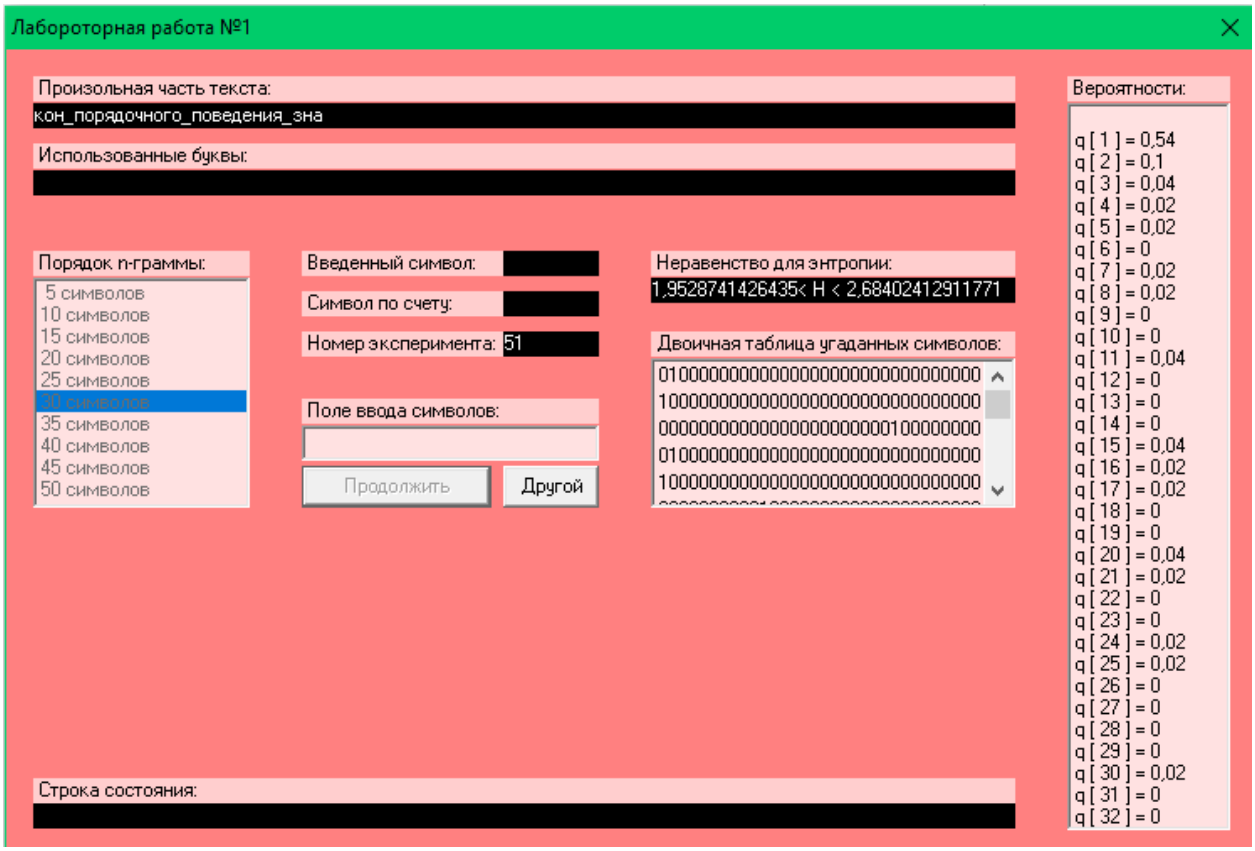
Ентропія біграм з пробілами: 3.9104802888234858

Надлишковість: 0.8777974909742661

Завдання 2


Лабороторная работа №1
Произольная часть текста:
лючением_
Использованные буквы:
Порядок n-граммы:
5 символов
10 символов
15 символов
20 символов
25 символов
30 символов
35 символов
40 символов
45 символов
50 символов
Введенный символ:
Символ по счету:
Номер эксперимента: 50
Поле ввода символов:
Продолжить
Другой
Неравенство для энтропии:
3,31379269302549< H < 3,85098771375663
Двоичная таблица угаданных символов:
Строка состояния:
Вероятности:
q [1] = 0,2857142
q [2] = 0,1020408
q [3] = 0,0408163
q [4] = 0,0204081
q [5] = 0,0408163
q [6] = 0,0204081
q [7] = 0,0204081
q [8] = 0,0204081
q [9] = 0,0612244
q [10] = 0
q [11] = 0,020408
q [12] = 0
q [13] = 0,020408
q [14] = 0
q [15] = 0,020408
q [16] = 0,040816
q [17] = 0,020408
q [18] = 0,040816
q [19] = 0,020408
q [20] = 0,040816
q [21] = 0
q [22] = 0
q [23] = 0
q [24] = 0
q [25] = 0
q [26] = 0,020408
q [27] = 0
q [28] = 0,020408
q [29] = 0
q [30] = 0,020408
q [31] = 0,040816
q [32] = 0,061224


Лабороторная работа №1
Произольная часть текста:
о_когда_мыслители_д
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:
2,10330345874846< H < 2,66111918856318
Двоичная таблица угаданных символов:
Строка состояния:
Вероятности:
q [1] = 0,54
q [2] = 0,06
q [3] = 0,08
q [4] = 0
q [5] = 0,02
q [6] = 0
q [7] = 0
q [8] = 0,02
q [9] = 0
q [10] = 0
q [11] = 0
q [12] = 0
q [13] = 0
q [14] = 0,04
q [15] = 0,02
q [16] = 0,02
q [17] = 0,04
q [18] = 0
q [19] = 0
q [20] = 0
q [21] = 0,04
q [22] = 0,02
q [23] = 0,02
q [24] = 0,02
q [25] = 0
q [26] = 0
q [27] = 0,02
q [28] = 0
q [29] = 0,04
q [30] = 0
q [31] = 0
q [32] = 0



Отримали значення:

3.31379 < H(10) < 3.85098

2.10330 < H(20) < 2.66119

1.95287 < H(30) < 2.68402

Значення надлишковості

0.33724 > R(10) > 0.22980

0.57934 > R(20) > 0.46776

0.60942 > R(30) >0.46319

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.