Міністерство освіти і науки України

Національний технічний університет України

“Київський політехнічний інститут ім. Ігоря Сікорського” Фізико-технічний інститут

КРИПТОГРАФІЯ КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконали студенти :

3го курсу групи ФБ-04

Музичка-Скрипка Олександра Тарасівна

Кузьмін Гліб Ігорович

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:

0. Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.

1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку 1H та 2H за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення 1H та 2H на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення 1H та 2H 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

Ми взяли текст з книги «Война и мир» 1ший том.

Результат:

Текст з пробілами:

Ентропія для монограм: 4.389513128998312

Надлишковість для монограм: 0.12209737420033773

Ентропія для біграм з перетином: 3.984551695044777

Надлишковість для біграм з перетином: 0.2030896609910464

Ентропія для біграм без перетину: 3.9839081215820165

Надлишковість для біграм без перетину: 0.20321837568359669

Текст без пробілів:

Ентропія для монограм: 4.46859115010279

Надлишковість для монограм: 0.09801895804289695

Ентропія для біграм з перетином: 4.149796060796719

Надлишковість для біграм з перетином: 0.16236745562618626

Ентропія для біграм без перетину: 4.150318718889217

Надлишковість для біграм без перетину: 0.16226195756762085

Монограми без пробілів

0	Кількість	Частота
о	61232	0,113683
а	45176	0,083874
е	42904	0,079656
и	35789	0,066446
н	35097	0,065161
т	30591	0,056795
с	28103	0,052176
л	27263	0,050616
в	24795	0,046034
р	24545	0,04557

Монограми з пробілами

0	Кількість	Частота
	102010	0,159234
о	61232	0,095581
а	45176	0,070518
е	42904	0,066972
и	35789	0,055865
н	35097	0,054785
т	30591	0,047752
с	28103	0,043868
л	27263	0,042557
в	24795	0,038704

Біграми:

З пробілами та з перетинами

0	Кількість	Частота
о	13207	0,020616
и	11287	0,017619
а	10404	0,01624
е	10309	0,016092
с	9791	0,015283
п	9648	0,01506
в	9451	0,014753
н	9221	0,014394
то	8490	0,013253
о	7533	0,011759

З пробілами та без перетинів

0	Кількість	Частота
о	6641	0,020733
и	5624	0,017558
а	5291	0,016518
е	5180	0,016172
с	4906	0,015316
п	4808	0,01501
в	4756	0,014848
н	4616	0,014411
то	4213	0,013153
о	3792	0,011838

Без пробілами та з перетинами

0	Кількість	Частота
то	8656	0,016071
ст	6836	0,012692
на	6582	0,01222
ов	6482	0,012035
ал	5855	0,01087
го	5799	0,010766
он	5618	0,01043
не	5579	0,010358
но	5541	0,010287
ко	5489	0,010191

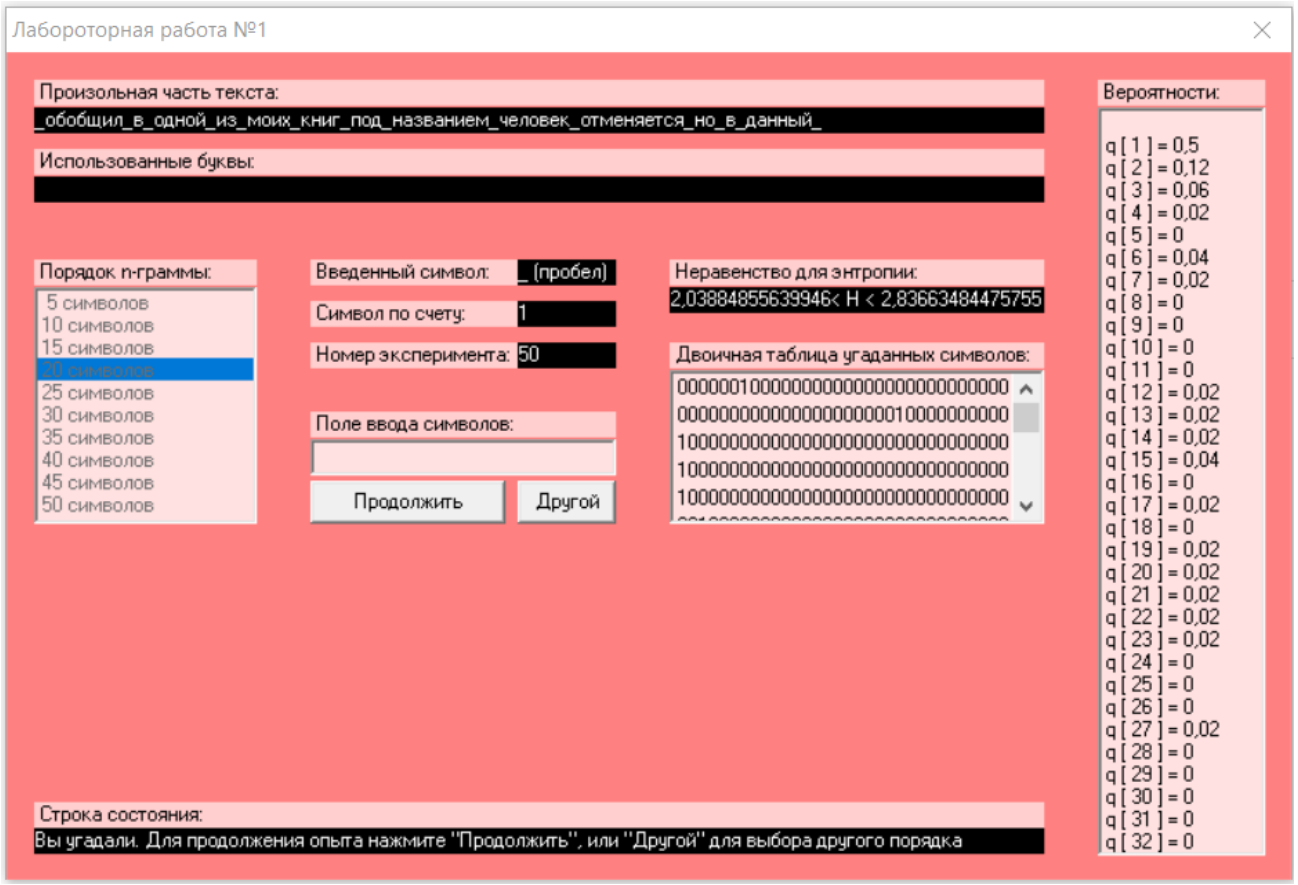
Без пробілами та без перетинів

0	Кількість	Частота
то	4322	0,016048
ст	3438	0,012766
ов	3251	0,012072
на	3236	0,012016
ал	2926	0,010865
го	2898	0,010761
не	2862	0,010627
он	2774	0,0103
ко	2757	0,010237
ос	2717	0,010089

Значення H(10):

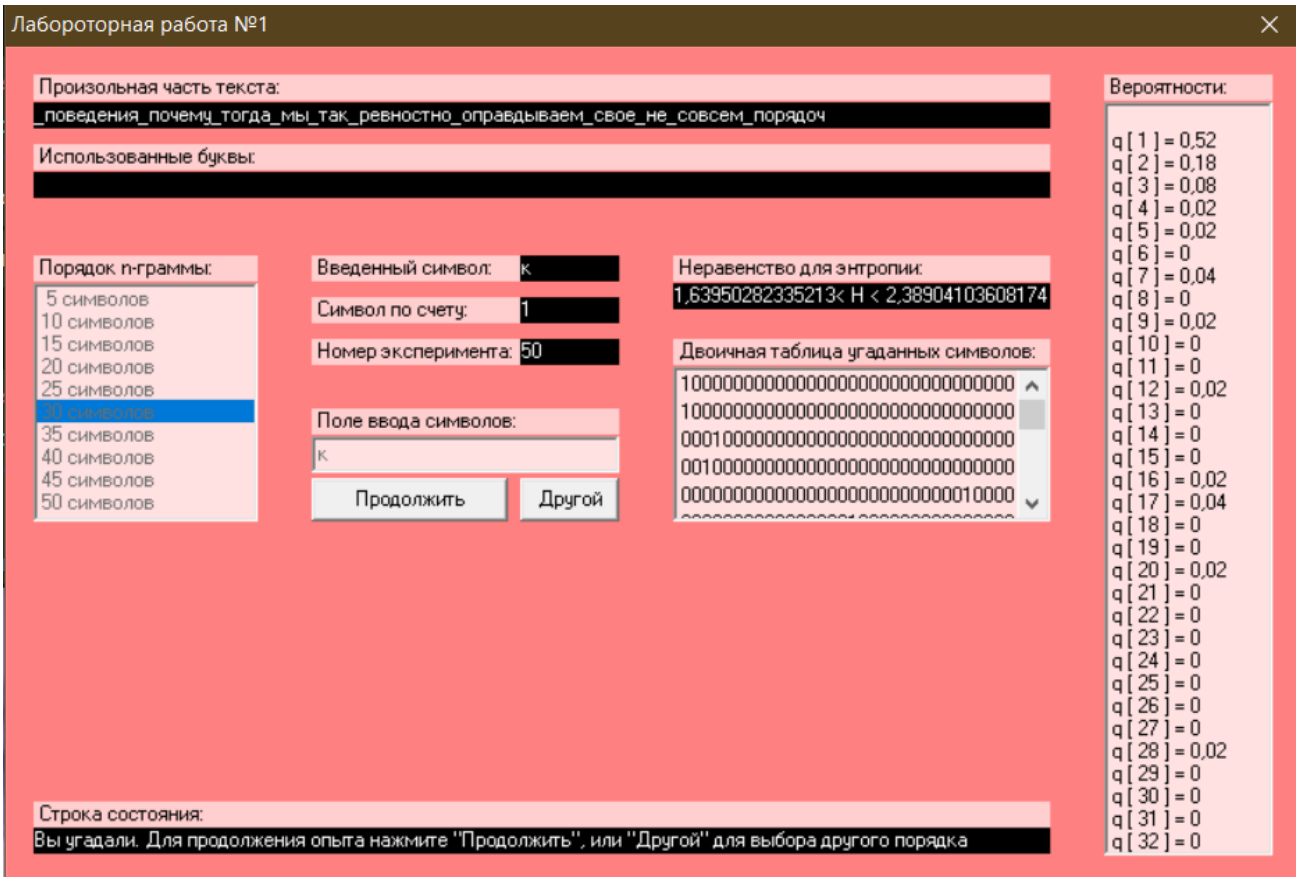
R = 0,523752290017084

Значення H(20):



R = 0,512451659884299

Значення H(30):



R = 0,597145614056613

Висновки:

В ході виконання лабораторної роботи ми закріпили теоретичний матеріал, а саме: що таке ентропія на символі джерела, надлишковість, біграми, дослідили на прикладі першого тому «Война и мир» наближене значення ентропії.

Під час аналізу тексті, ми підтвердили факт, що в російському алфавіті найчастіше зустрічається пробіл. З цього можна зробити висновки, що при шифрувані слід уникати цього символу, для ускладнення зламу шифрування