Міністерство освіти і науки України

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Фізико-технічний інститут

ЛАБОРАТОРНА РОБОТА №1

з дисципліни

«Криптографія»

на тему: «Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Виконали:

студенти 3 курсу ФТІ

групи ФБ-74,ФБ-72

Демиденко Дар’я та Скуратов Ілля

Перевірили:

Чорний О.

Савчук М. М.

Завадська Л. О.

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 H_1 та H_2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H_1 та H_2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 на тому ж тексті, в якому вилучено всі пробіли.

2. За допомогою програми CoolPinkProgram оцінити значення H^{10} , H^{20} , H^{30} .

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

Код програми

Результати виконнання програми:

Монограми:

| н - 0.057492повторение: 57322 -
| е - 0.0712146повторение: 71004 -
| л - 0.023688повторение: 23200 -
| р - 0.0356484повторение: 35543 -
| и - 0.0558822повторение: 55717 -
| м - 0.0285524повторение: 28468 -
| ы - 0.0150024повторение: 14958 -
| й - 0.00868568повторение: 8660 -
| - 0.161101повторение: 160625 -
| с - 0.0445919повторение: 44460 -
| г - 0.0143374повторение: 14295 -
| б - 0.0130827повторение: 13044 -
| а - 0.0712627повторение: 71052 -
| ж - 0.00910793повторение: 9081 -
| т - 0.053416повторение: 53258 -
| о - 0.09199повторение: 91718 -
| к - 0.0253219повторение: 25247 -
| д - 0.0409651повторение: 40844 -
| д - 0.0260841повторение: 26007 -
| в - 0.035832повторение: 35726 -
| з - 0.0138149повторение: 13774 -
| у - 0.0228345повторение: 22767 -
| ч - 0.012865повторение: 12827 -
| ь - 0.0181988повторение: 18145 -
| я - 0.0204164повторение: 20356 -
| ш - 0.00674795повторение: 6728 -
| х - 0.0061552повторение: 6137 -
| щ - 0.00301592повторение: 3007 -
| ю - 0.00536085повторение: 5345 -
| ц - 0.0031483повторение: 3139 -
| ф - 0.000854527повторение: 852 -
| э - 0.00262777повторение: 2620 -
| ъ - 0.000185549повторение: 185 -
| ё - 0.000934764повторение: 932 -

Количество символов: 836418
******

| н - 0.0685327повторение: 57322 -
| е - 0.0848906повторение: 71004 -
| л - 0.0277373повторение: 23200 -
| р - 0.0424943повторение: 35543 -
| и - 0.0666138повторение: 55717 -
| м - 0.0340356повторение: 28468 -
| ы - 0.0178834повторение: 14958 -
| й - 0.0103537повторение: 8660 -
| с - 0.0531552повторение: 44460 -
| г - 0.0170907повторение: 14295 -
| б - 0.0155951повторение: 13044 -
| а - 0.084948повторение: 71052 -
| ж - 0.010857повторение: 9081 -
| т - 0.0636739повторение: 53258 -
| о - 0.109656повторение: 91718 -
| к - 0.0301847повторение: 25247 -
| д - 0.048832повторение: 40844 -
| д - 0.0310933повторение: 26007 -
| в - 0.0427131повторение: 35726 -
| з - 0.016467Bповторение: 13774 -
| у - 0.0272196повторение: 22767 -
| ч - 0.0153356повторение: 12827 -
| ь - 0.0216937повторение: 18145 -
| я - 0.0243371повторение: 20356 -
| ш - 0.00804382повторение: 6728 -
| х - 0.00733724повторение: 6137 -
| щ - 0.00359509повторение: 3007 -

З пробілом

H1 = 4.37681

Без пробіла

H1 = 4.45808

Біграми:

Перетинаються, з пробілом – H2 = 3.98984

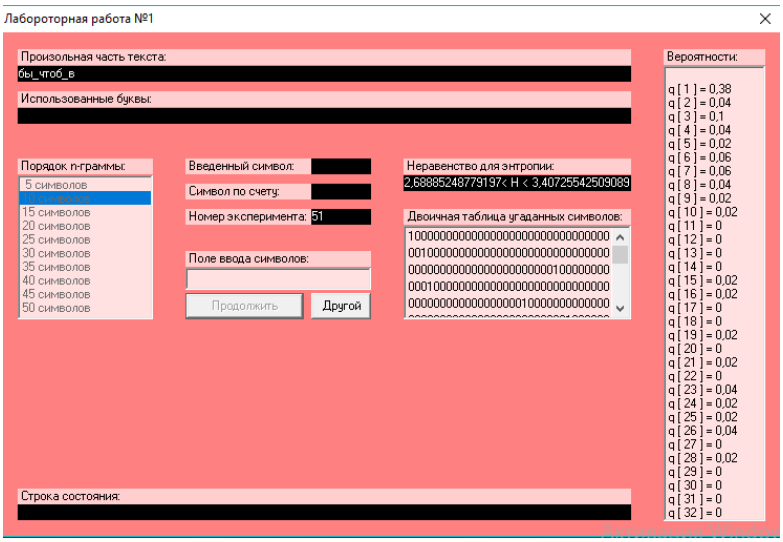
Не перетинаються, з пробілом – H2 = 3.98924

Перетинаються, без пробіла – H2 = 4,14112

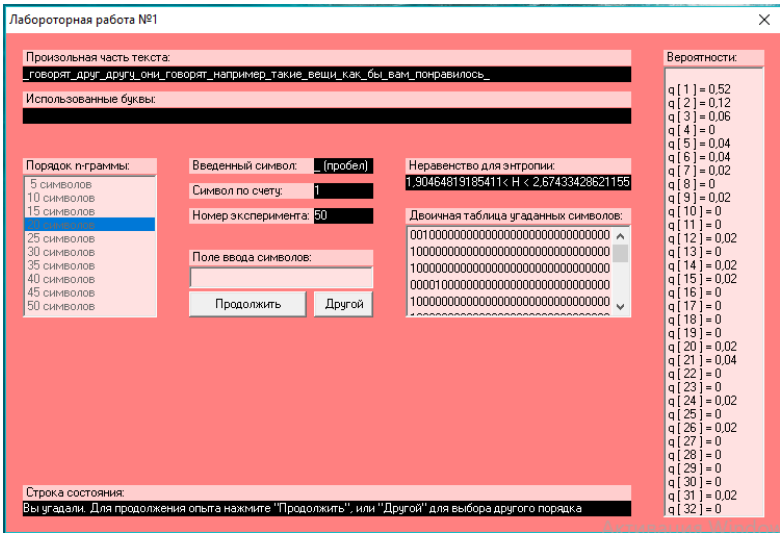
Не перетинаються, без пробіла – H2 = 4.14115

2.Скріншоти результатів рожевої программи:

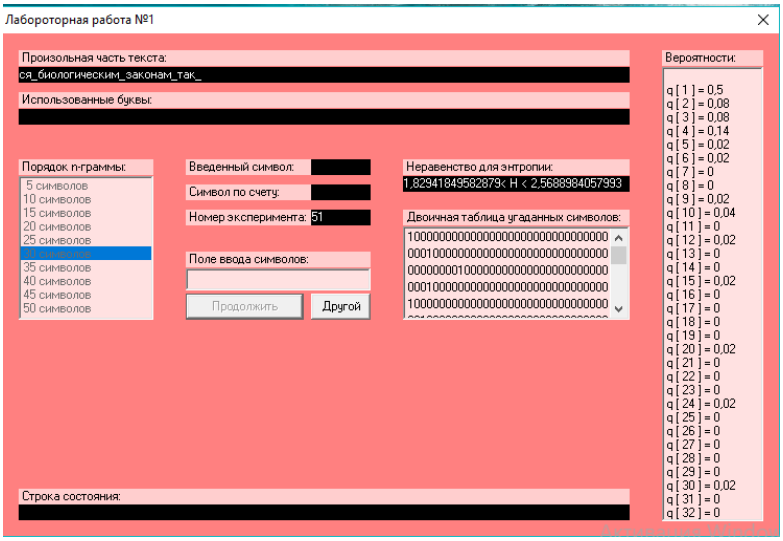
10-грамми:



20-грамми:



30-грами:



3. Оцінка надлишковості російської мови в різних моделях джерела.

З пробілом, H1: R1=1-H1/ H0=0.124638

Без пробіла, H1: R1=1-H1/ H0=0.108384

Перетинаються, з пробілом, H2: R2=1-H2/ H0= 0.202032

Не перетинаються, з пробілом, H2: R2=1-H2/ H0=0.202152

Перетинаються, без пробіла, H2: R1=1-H2/ H0=0.171776

Не перетинаються, без пробіла, H2 R1=1-H2/ H0=0.17177

Висновки:

Під час данного комп’ютерного практикуму, ми навчились визначати ентропію на символ джерела та його надлишковості. Порівняли різні моделі джерел відркритого тексту для наближенного визначення ентропії та набули практичних навичок оцінки ентропії на символ джерела.