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:

Дудченко И.В і Терпило С.Е

Київ – 2022

***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***

Уважно прочитати методичні вказівки до виконання комп’ютерного практикуму.

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

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

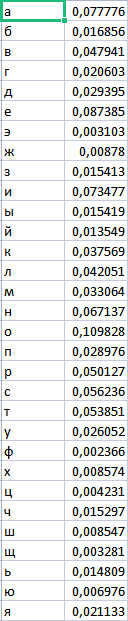
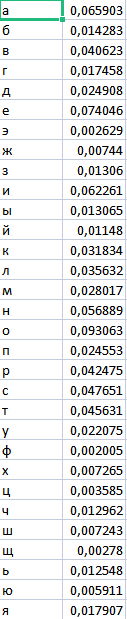
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

***Хід роботи***

Перед виконанням роботи були розглянуті теоретичні відомості в методичних вказівках. В якості експериментального тексту була взята книга «Good Omens» в перекладі на російську мову. Оригінал тексту можна знайти у файлі badtxt.txt. Відредагований текст з пробілами міститься у файлі spaces.txt, а без пробілів у nospaces.txt. В ході виконання роботи було прийнято рішення пробіли замінити на «\_» для кращого сприйняття. Усі таблиці також наведені у файлах з відповідними назвами. Єдина відмінність - таблиці біграм були виконанні у двох варіантах(таблиці у вигляді матриці і звичайні*).*

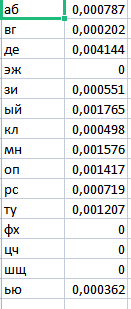
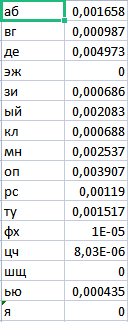
**Аналіз результатів для тексту з пробілами**

**Текст без пробілів Текст з пробілами**

** **

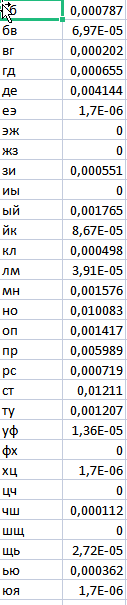
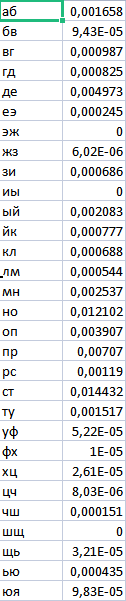
**Біграми**

**Текст з пробілами Текст без пробілів**

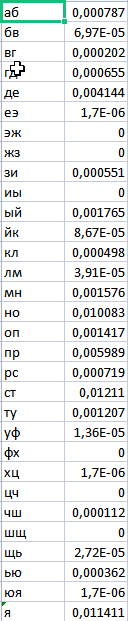
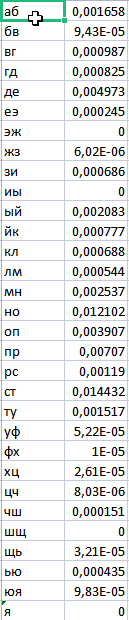
** **

**Перехрестні біграми**

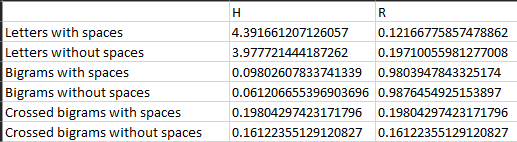
**crossed\_bigrams(text\_with\_space) crossed\_bigrams(text\_without\_space)**

** **

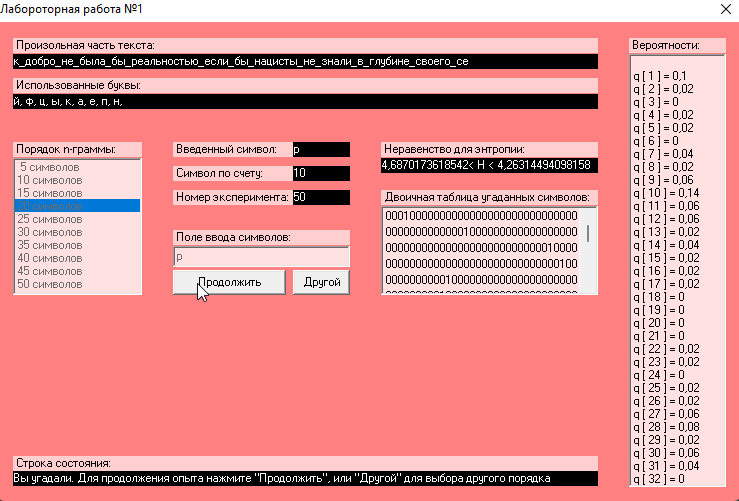
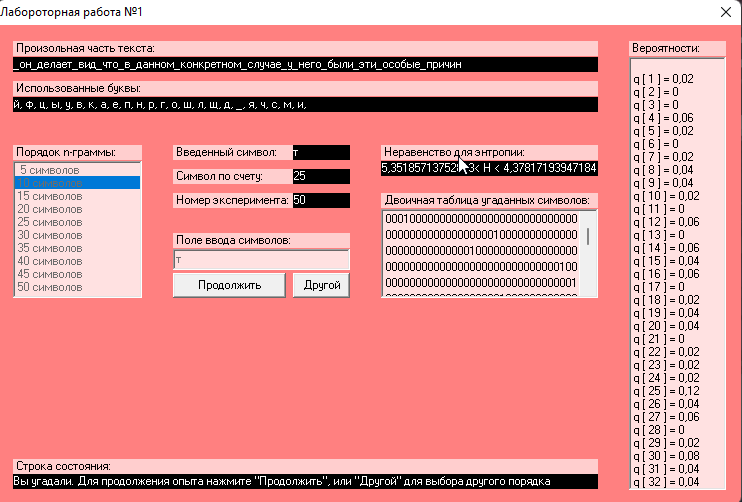
**crossed\_bigrams\_space(text\_with\_space) crossed\_bigrams\_space(text\_without\_space)**

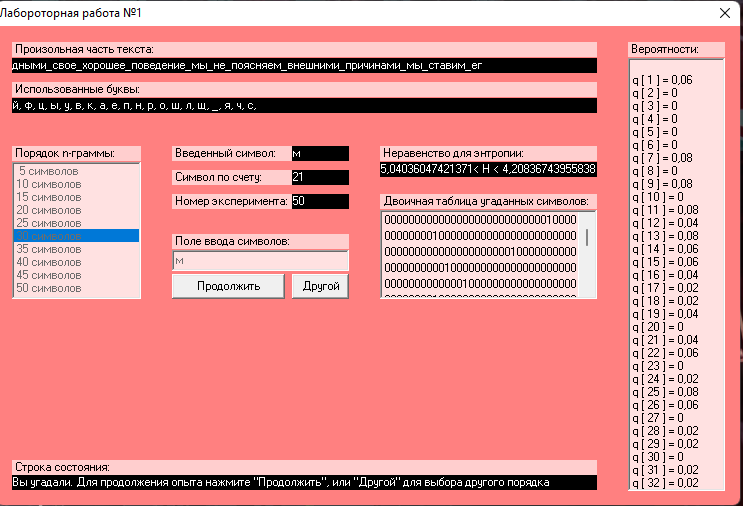
** **

**Значення H і R**

****

**CoolPinkProgram**

****

****

**Отримані результати**

5,3518<H(10)<4,3781

4,6870<H(20)<4,2631

5,0403<H(30)<4,2083

**Висновки**

Під час виконання лабораторної роботи, мы отрамали змогу ознайомитись з такими поняттями як ентропія, надлишковість та обрахувати їх на практиці. Успішно були проведені експерименти на різних видах тексту(з пробілом та без), порахована частота монограм та біграм, проведено знайомство з невеличкою програмою. Набуті навички знадобляться у майбутніх лаб.роботах та у професійній діяльності.