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-72

Макоїд Ігор та Оліферук Артур

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 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 H10 , H20 , H30 .

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

**Код програми:**

**Файл test.cpp містить в собі код для першочергової обробки і заміни абзаців, відступівЮ кількох пробілів і так на далі на один символ пробілу. Також ми використовували даний код для видалення пробілів взагалі коли це потрібно.**

**test.cpp**

**#include <fstream>**

**#include <iostream>**

**#include <string>**

**using namespace std;**

**int main()**

**{**

**setlocale(LC\_ALL, "RUS");**

**wifstream in("rinput.txt");**

**wofstream out("output.txt");**

**wstring t,r;**

**int i =0;**

**while(1)**

**{**

**in>>t;**

**if(in.eof())**

**break;**

**out<<t;**

**}**

**}**

**Код який міститься у файлі Reader.cpp ми використовували для подальшої обробки тексту а саме заміни ‘ё’ на ‘е’, ‘ъ’ на ‘ь’ а також усіх великих літер на маленькі.**

**Reader.cpp**

**#include <fstream>**

**#include <iostream>**

**#include <string>**

**using namespace std;**

**int main()**

**{**

**setlocale(LC\_ALL, "RUS");**

**wifstream in("rinput.txt");**

**wofstream out("output.txt");**

**wstring t,r;**

**int i =0;**

**while(1)**

**{**

**in>>t;**

**for(int s=0;s<t.length();s++)**

**{**

**if((int)t[s]==1025)**

**t[s]=1077;**

**if((int)t[s]==1105)**

**t[s]=1077;**

**if((int)t[s]==1066)**

**t[s]=1100;**

**if((int)t[s]==1098)**

**t[s]=1100;**

**if(((int)t[s]>=1040)and((int)t[s]<=1071 ))**

**{**

**t[s]+=32;**

**}**

**i++;**

**}**

**if(in.eof())**

**break;**

**out<<t<<" ";**

**}**

**}**

**Код файлу Filter.cpp ми використовували для остаточної фільтрації вхідного і тексту і вилучення з нього посторонніх символів по типу “,.!” і так далі.**

**Filter.cpp**

**#include <fstream>**

**#include <iostream>**

**#include <string>**

**using namespace std;**

**int main()**

**{**

**setlocale(LC\_ALL, "RUS");**

**wifstream in("1.txt");**

**wofstream out("output.txt");**

**wstring t,r;**

**int i =0;**

**while(1)**

**{**

**in>>t;**

**for(int s=0;s<t.length();s++)**

**{**

**if(!(t[s]>=1072 and t[s]<=1103) )**

**t[s]=' ';**

**}**

**out<<t<<" ";**

**i++;**

**if(in.eof())**

**break;**

**}**

**}**

**Код файлу Monograms.cpp використовується нами для обрахування ентропії монограм з пробілами та без них.**

**Monograms.cpp**

**#include <fstream>**

**#include <iostream>**

**#include <string>**

**#include <math.h>**

**using namespace std;**

**int main()**

**{**

**setlocale(LC\_ALL, "RUS");**

**wifstream in("input.txt");**

**wofstream out("output.txt");**

**wstring t,r;**

**int alfa[33];**

**for(int letter=0;letter<32;letter++)**

**{**

**alfa[letter]=0;**

**alfa[32]=213495;**

**}**

**while(1)**

**{**

**in>>t;**

**for(int i=0;i<t.length();i++)**

**{**

**int a =1072;**

**while(a!=1104)**

**{**

**if((int)t[i]==a)**

**{**

**alfa[a-1072]+=1;**

**break;**

**}**

**else**

**a++;**

**}**

**}**

**if(in.eof())**

**break;**

**}**

**int sumlet=0;**

**for(int letnum=224, letter=0;letnum<256;letnum++, letter++)**

**{**

**if(letnum==250)**

**continue;**

**cout<<(char)letnum<<" "<<alfa[letter]<<endl;**

**sumlet+=alfa[letter];**

**}**

**cout<<"\_ "<< alfa[32]<<endl;**

**cout<<"sumlet = "<<sumlet+alfa[32]<<endl;**

**cout<<"=============================================="<<endl;**

**cout<<"LETTER FREQUENCY + \_"<<endl;**

**double H = 0.0;**

**for(int letnum=224, letter=0;letter<33;letnum++, letter++)**

**{**

**if(letnum==250)**

**continue;**

**double pi = (double)alfa[letter]/(double)(sumlet+alfa[32]);**

**cout<<(char)letnum<<" "<<pi<<endl;**

**double l = log10(pi)/log10(2);**

**H=H+(pi\*l);**

**}**

**cout<<"=============================================="<<endl;**

**cout<<"LETTER FREQUENCY"<<endl;**

**double h=0.0;**

**for(int letnum=224, letter=0;letter<32;letnum++, letter++)**

**{**

**if(letnum==250)**

**continue;**

**double pi = (double)alfa[letter]/(double)sumlet;**

**cout<<(char)letnum<<" "<<pi<<endl;**

**double l = log10(pi)/log10(2);**

**h=h+(pi\*l);**

**}**

**cout<<"sumlet = "<<sumlet<<endl;**

**cout<<"=============================================="<<endl;**

**cout<<"Entropy H = "<<H\*(-1)<<endl;**

**cout<<"Entropy h = "<<h\*(-1);**

**}**

**Код файлу Creatorbi.cpp використовується для створення алфавіту усіх можливих біграм з пробілами чи без залежно від вхідних файлів.**

**Creatorbi.cpp**

**#include <fstream>**

**#include <iostream>**

**#include <string>**

**using namespace std;**

**int main()**

**{**

**setlocale(LC\_ALL, "RUS");**

**wifstream in("3.txt");**

**wofstream out("output.txt");**

**wstring t,r;**

**int i =0, e=0;**

**while(1)**

**{**

**in>>t;**

**e++;**

**wcout<<t<<endl;**

**wifstream in2("4.txt");**

**while(1)**

**{**

**in2>>r;**

**out<<t+r<<" ";**

**i++;**

**if(in2.eof())**

**break;**

**}**

**in2.close();**

**if(in.eof())**

**break;**

**}**

**}**

**Код файлу Entropiabitekstzpr.cpp ми використовували для обрахування ентропії для біграм в тексті у випадку якщо пробіл входить в наш алфавіт.**

**Entropiabitekstzpr.cpp**

**#include <fstream>**

**#include <iostream>**

**#include <string>**

**#include <math.h>**

**using namespace std;**

**int main()**

**{**

**setlocale(LC\_ALL, "RUS");**

**wifstream in("vsebigramizprobelami.txt");**

**wofstream out("output.txt");**

**wifstream in2("tekstzprobelami.txt");**

**wstring t,r;**

**int i = 0, e = 0;**

**double f=0.0, entropia = 0.0,ren,logar=0.0;**

**while(1)**

**{**

**in>>t;**

**in2>>r;**

**for(int s=0;s<r.length(); s++)**

**{**

**if(t[0]==r[s] and t[1]==r[s+1])**

**e++;**

**}**

**out<<t<<" - "<<e<<endl;**

**cout<<i<<endl;**

**if(e!=0)**

**{**

**f = (double)e/1233599;**

**// f = (double)e/616800;**

**cout<<"×àñòîòà: "<<f<<" ";**

**logar=log10(f)/log10(2);**

**cout<<"Ëîãàðèôì :"<<logar<<endl;**

**ren=f\*logar;**

**entropia+=ren;**

**}**

**e=0;**

**i++;**

**if(in.eof())**

**break;**

**}**

**entropia = entropia-ren;**

**cout<<"À åíòðîïèÿ ó íàñ âîò òàêà: "<<(entropia\*-(1))/2;**

**}**

**Код файлу Еntropiabibezpr.cpp ми використовували для обрахунку значення ентропії в випадку коли у тексті не враховуються пробіли.**

**Entropiabibezpr.cpp**

**#include <fstream>**

**#include <iostream>**

**#include <string>**

**#include <math.h>**

**using namespace std;**

**int main()**

**{**

**setlocale(LC\_ALL, "RUS");**

**wifstream in("vsebigamibezprobelov.txt");**

**wofstream out("output.txt");**

**wifstream in2("tekstbezprobelov.txt");**

**wstring t,r;**

**int i = 0, e = 0;**

**double f=0.0, entropia = 0.0,ren,logar=0.0;**

**while(1)**

**{**

**in>>t;**

**in2>>r;**

**for(int s=0;s<r.length(); s+=2)**

**{**

**if(t[0]==r[s] and t[1]==r[s+1])**

**e++;**

**}**

**out<<e<<endl;**

**cout<<i<<endl;**

**if(e!=0)**

**{**

**//f = (double)e/1024417;**

**f= (double)e/512209;**

**cout<<"×àñòîòà: "<<f<<" ";**

**logar=log10(f)/log10(2);**

**cout<<"Ëîãàðèôì :"<<logar<<endl;**

**ren=f\*logar;**

**entropia+=ren;**

**}**

**e=0;**

**i++;**

**if(in.eof())**

**break;**

**}**

**entropia = entropia-ren;**

**cout<<"À åíòðîïèÿ ó íàñ âîò òàêà: "<<(entropia\*-(1))/2;**

**}**

**Результати виконнання програми:**

Надлишковість та ентропія монограм тексту з пробілами

H = 4.35009, R = 0.129982

Надлишковість та ентропія монограм тескту без пробілів

H = 4.45513, R = 0.108974

Надлишковість та ентропія біграм тексту з пробілами без перетину

H = 3.93864, R = 0.212272

Надлишковість та ентропія біграм тексту без пробілів, що перетинаються

H = 4.12719, R = 0.174562

Надлишковість та ентропія біграм тексту з пробілами, що перетинаються

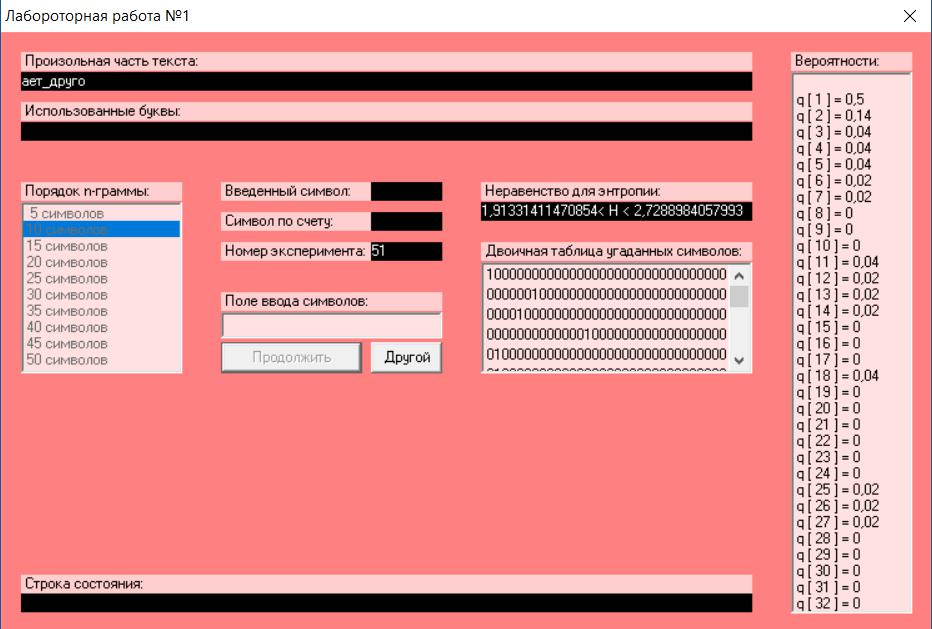
H = 3.9389, R = 0.21222

Надлишковість та ентропія біграм тексту без пробілів без перетину

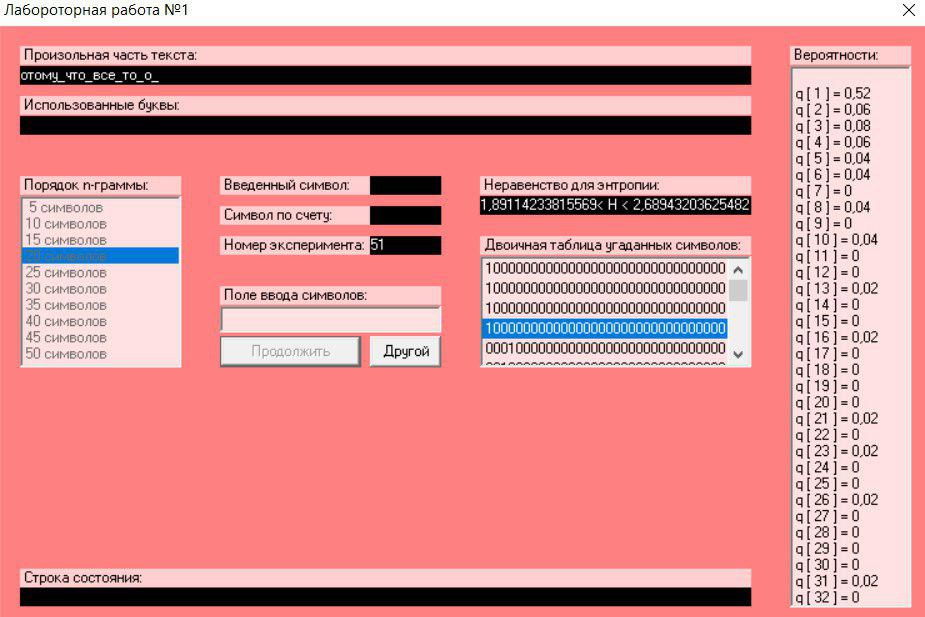
H = 4.12617, R = 0.174766

**Скріншоти результатів рожевої программи:**

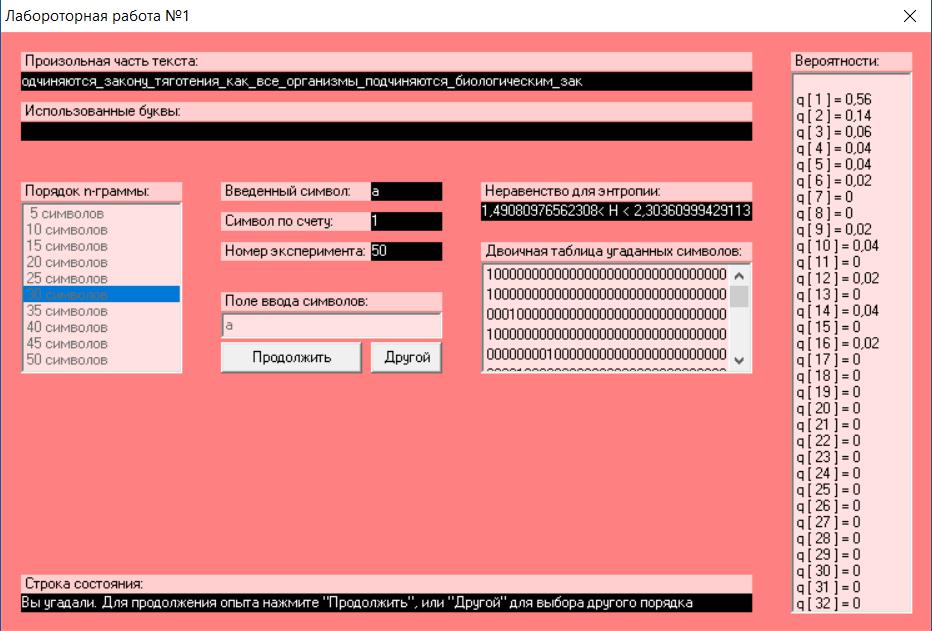
**Н10:**

****

**Н20:**

****

**Н30:**

****

**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.