Міністерство освіти і науки України

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Фізико-технічний інститут

**ЛАБОРАТОРНА РОБОТА №1**

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

**Виконали:**

студенти 3 курсу ФТІ

групи ФБ-73

Шишкін М.В. та Вітрович А.К.

Перевірив:

Чорний 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**

from math import log

a="абвгдеёжзийклмнопрстуфхцчшщъыьэюя "

file = open("1.txt", 'r',encoding='utf-8')

file\_2 = open("2.txt",'w')

line = file.readlines()

gram={}

bigram={}

bigram\_u={}

len1=0

len2=0

len3=0

def chek(aa,q):

for l in line:

for elem in l.lower():

if elem not in aa:

l=l.replace(elem, ' ')

l = ' '.join(l.split())

if q==1:

l = l.replace(' ', '')

file\_2.write(l.lower() + '')

chek(a,1)

file\_3 = open("2.txt",'r')

line = file\_3.readlines()

def gramma():

len1=0;

for l in line:

for elem in l.lower():

if elem in gram:

len1 = len1 + 1

gram[elem]=gram[elem]+1;

else:

gram[elem]=1;

len1 = len1 + 1

return len1;

len1 = gramma()

file\_3.close()

print (gram)

file\_4 = open("2.txt",'r')

line = file\_4.readlines()

def bigramma():

a=''

len2=0;

for l in line:

for elem in l.lower():

t=elem+a

if len(t)>1:

if t in bigram:

bigram[t]=bigram[t]+1;

len2 = len2 + 1

else:

bigram[t]=1;

len2=len2+1

a = elem

return len2;

len2 = bigramma()

print(bigram)

file\_4.close()

file\_5 = open("2.txt",'r')

line = file\_5.readlines()

def bigramma\_u():

a=''

len2=0;

for l in line:

for i in range(1,len(l)):

t=l[i]+a

if len(t)>1:

if t in bigram\_u:

bigram\_u[t]=bigram\_u[t]+1;

len2 = len2 + 1

else:

bigram\_u[t]=1;

len2=len2+1

a = l[i]

i=i+2

return len2;

len3 = bigramma\_u()

print(bigram\_u)

file\_5.close()

#/////////////

e=0;

for value in gram.keys():

v=gram[value] /len1;

e -= v \* log(v,2)

print("Entropia in leler:")

print(e)

print("RESIDUALITY in leler")

print(1-e/log(len(a),2))

e=0;

for value in bigram.keys():

v=bigram[value] /len2;

e -= v \* log(v,2)

e=e/2

print("Entropia in bigram:")

print(e)

print("RESIDUALITY in bigram:")

print(1-e/log(len(a),2))

e=0;

for value in bigram\_u.keys():

v=bigram\_u[value] /len2;

e -= v \* log(v,2)

e=e/2

print("Entropia in bigram(without intersections):")

print(e)

print("RESIDUALITY in bigram(without intersections):")

print(1-e/log(len(a),2))

file.close()

file\_2.close()

**Результати виконнання програми:**

**З пробілом:**

Entropia in leler:

4.379492516972483

RESIDUALITY in leler:

0.1391598025124562

Entropia in bigram:

3.9536470619319046

RESIDUALITY in bigram:

0.2228646802341614

Entropia in bigram(without intersections):

3.9536391207928236

RESIDUALITY in bigram(without intersections):

0.22286624115742093

**Без пробілу:**

Entropia in leler:

4.379492516972483

RESIDUALITY in leler

0.1391598025124562

Entropia in bigram:

3.9536470619319046

RESIDUALITY in bigram:

0.2228646802341614

Entropia in bigram(without intersections):

3.9536391207928236

RESIDUALITY in bigram(without intersections):

0.22286624115742093

**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.