Изображение выглядит как клавиатура, рисунок

Автоматически созданное описаниеМіністерство освіти і науки України

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Фізико-технічний інститут

**ЛАБОРАТОРНА РОБОТА №1**

З дисципліни«Криптографія»

**Виконали:**

студенти 3 курсу ФТІ

групи ФБ-93

Абдуллаєва Есміра

Шовак Мирослав

**Викладач:**

Селюх П. В.

м.Київ – 2021

**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.

**Завдання**

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 H (10), H (20), H (30).
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

**Хід роботи та опис труднощів:**

Під час роботи , з огляду на вимоги форматування тексту , та різних операцій над ним , було вирішено використовувати мову python. У якості вхідного тексту була обрана книга “Зелена миля”. Форматування вхідного тексту не викликало значних труднодщів . За допомогою стандартних функцій ми змогли позбутися пробілів , твердого знаку , змінити всі великі букви на малі. За допомогою регулярного виразу змогли досягнути зчитування з файлу тільки алфавіту. Для зручнішого використання програми , в залежності від завдання , було вирішено зробити лаконічне меню з 4 можливостями : зчитування файлу з пробілами і виконання операцій над знаками , зчитування файлу без пробілів , аналогічно було проведено для біграм. Сортування монограм ми досягли за допомогою numpy. Труднощі виникли при створення функції для зчитування біграм. Було прийнято рішення створити масив букв , і вже до значень цього масиву додавати ,як аргумент фукції ще одну букву по черзі і потім робити зчитування циклічно для кожної отриманої пари. Вивід реалізовували у вигляді квадратної матриці , для презентабельного вигляду ми відформатували значення наших H2 до трьох знаків після коми.

**Код програми**

import math  
import re  
import numpy as np  
file = open("/Users/esmira.23/Desktop/КПИ/3курс/Крипта/1.txt", "r")  
data = file.read().replace("ъ","ь")  
one = re.sub(r"[^а-я]+", " ", data.lower()).replace("ъ","ь").replace("ё","е")  
  
*# m=32 - кількість букв алфавіту*alphabet = ['а','б','в','г','д','е','ж','з','и','й','к','л','м','н','о','п','р','с','т','у','ф','х','ц','ч','ш','щ','ы','ь','э','ю','я',' ']  
save = []  
save1 = []  
save2 = []  
save3 = []  
save4 = []  
  
print("1. First task\n2. Second task\n3. Third task\n4. Fourth task\n")  
ans = input()  
  
def func():  
 for i in alphabet:  
 letter = one.count(i)  
 if letter == 0:  
 y = 0  
 else:  
 y = -math.log2(letter/number)  
 H1 = letter/number \* y  
 save.append(H1)  
 save1.append(i)  
 save4.append("%.4f" % H1)  
 a = np.column\_stack([save1,save4])  
 b = sorted(a, key = lambda x: x[1])  
 c = "\n".join(map(str,b))  
 print(c)  
 print("H1: ", (sum(save)))  
  
def bigram(a):  
 for i in alphabet:  
 r = [alphabet[a] + i]  
 for j in r:  
 letter = one.count(j)  
 if letter == 0:  
 y1 = 0  
 else:  
 y1 = -math.log2(letter/number)  
 H2 = letter/number \* y1  
 save1.append("%.3f" % H2)  
 save3.append(H2)  
 save2.append(i)  
 print(alphabet[a], " ".join(map(str, save1)))  
 save1.clear()  
  
if ans == "1":  
 number = len(one)  
 func()  
elif ans == "2":  
 one = one.replace(" ", "")  
 number = len(one)  
 func()  
elif ans == "3":  
 number = len(one)  
 arr = [0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,24,26,27,28,29,30,31]  
 print(" ", " ".join(alphabet))  
 for item in arr:  
 bigram(item)  
 print("H2: ", (sum(save3)/2))  
elif ans == "4":  
 one = one.replace(" ", "")  
 number = len(one)  
 arr = [0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,24,26,27,28,29,30,31]  
 print(" ", " ".join(alphabet))  
 for item in arr:  
 bigram(item)  
 print("H2: ", (sum(save3)/2))  
else:  
 print("Wrong input")

Значення Н1 для тексту з пробілами:

Изображение выглядит как стол

Автоматически созданное описание

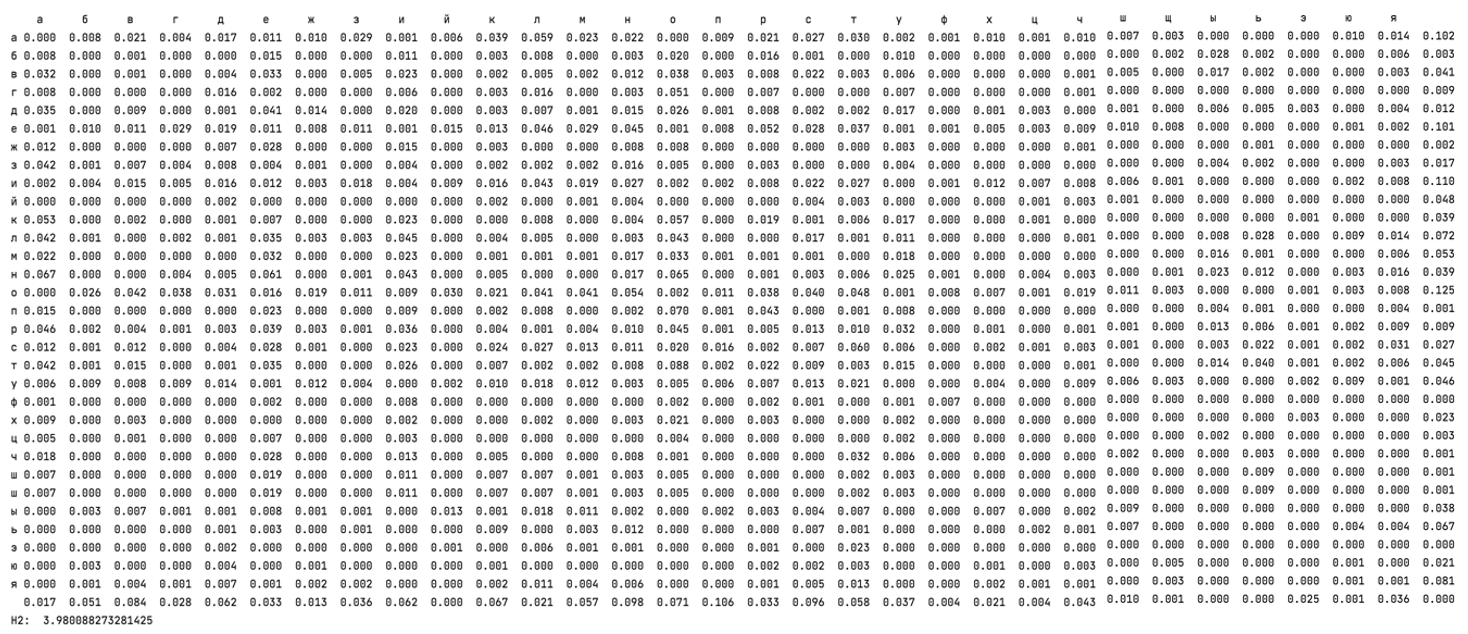
Значення Н1 для тексту без пробілів:

Изображение выглядит как стол

Автоматически созданное описание

Значення Н2 з пробілами:

H2: 3.980088273281425



Значення Н2 без пробілів:

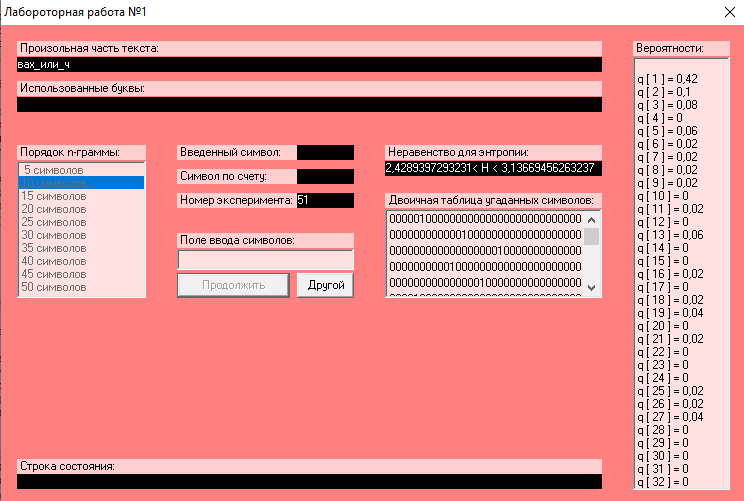
H2: 4.1878402723119

Изображение выглядит как текст

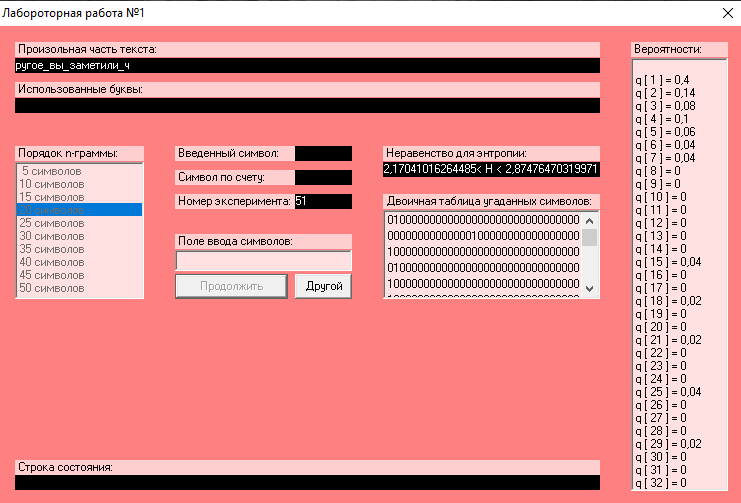
Автоматически созданное описание

За допомогою програми CoolPinkProgram оцінили значення  . У кожному досліді було зроблено не менше ніж 50 експериментів.

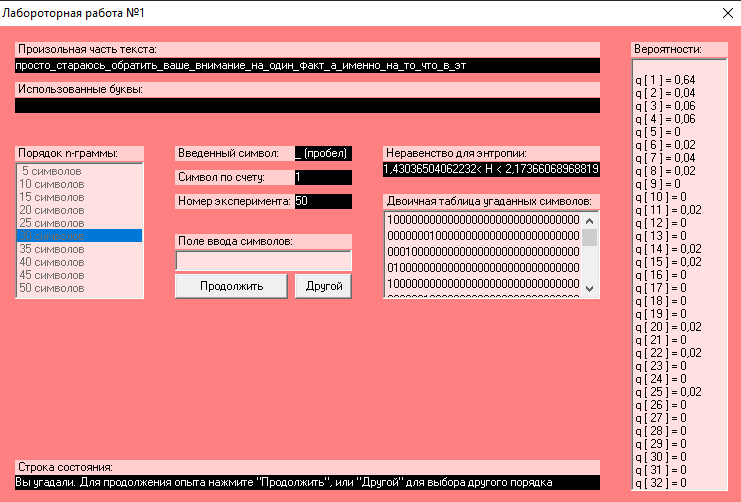
**Результат для H(10)**



**Результат для H(20)**



**Результат для H(30)**



**Результат надлишковості російської мови**

****

**Н0=log232=5**

**1-(2.428/5)=0.5144>R(10)>1-(3.136/5)=0.3728**

Аналогічні дії повели і для інших випадків.

**0.566>R(20)>0.4252**

**0.714>(30)>0.5654**

**Висновки:** під час лабораторної роботи , ми закріпили свої знання стосовно етропії та надлишковості тексту. Вирахували значення для вхідного тексту для монограм та біграм. Удосконалили свої знання мови python для роботи з різними структурами даних та файлами, навчилися краще придумовувати алгоритми для роботи з вхідним текстом.