**МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ**

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**

**«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

**ФІЗИКО- ТЕХНІЧНИЙ ІНСТИТУТ**

# Кафедра інформаційної безпеки

КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1

з дисципліни

Криптографія

З теми: « Експериментальна оцінка ентропії на символ джерела відкритого тексту »

Перевірила: Виконав студент групи ФБ-92

Селюх П.В Ханас Максим Любомирович

**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. Уважно прочитати методичні вказівки до виконання комп’ютерного практикуму.
2. 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.
3. За допомогою програми CoolPinkProgram оцінити значення H(10) , H(20), H(30).
4. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела

**1.**

Код програми знаходиться у файлах **main.py**

**Результати:**

H1 with whitespaces: 4.348237624893995

H2 with whitespaces: 3.935781747373973

H1 without whitespaces: 4.448697082249102

H2 without whitespaces: 4.125580893321075

Character frequency without whitespaces

|  |
| --- |
| о 0.114725160004508 |
| е 0.0871670156415899 |
| а 0.07966000154221213 |
| н 0.06508609695769 |
| и 0.06484764723677108 |
| т 0.06475155555819181 |
| с 0.05293109277592251 |
| в 0.0462616184923097 |
| л 0.04596148028637693 |
| р 0.04183191074150745 |
| к 0.03302706582279983 |
| д 0.032017510039207775 |
| м 0.03143977365071268 |
| у 0.02965199390233052 |
| п 0.027441885295007386 |
| ь 0.023226900924734117 |
| я 0.02136438320412364 |
| ч 0.018103197717526054 |
| б 0.017389034871788788 |
| г 0.016890781723599996 |
| ы 0.016513532911399915 |
| з 0.015397208596053123 |
| ж 0.011408810776503806 |
| й 0.010013701961575193 |
| х 0.008508265663833347 |
| ш 0.00823066748127102 |
| ю 0.005617211087318864 |
| э 0.0035269204989649386 |
| щ 0.00299070520615224 |
| ц 0.00277242287456477 |
| ф 0.0012444465534524791 |
| 0.0 |

Character frequency with whitespaces

|  |
| --- |
| 0.17087078018969903 |
| о 0.09512198240714967 |
| е 0.07227271967210373 |
| а 0.06604843492878171 |
| н 0.053964784791027114 |
| и 0.053767079159957626 |
| т 0.053687406741466936 |
| с 0.04388671565700729 |
| в 0.03835685964769053 |
| л 0.038108006291170234 |
| р 0.03468405951628022 |
| к 0.027383705318281475 |
| д 0.02654665311907682 |
| м 0.02606763499802786 |
| у 0.024585334570059104 |
| п 0.022752868944773245 |
| ь 0.019258102242335956 |
| я 0.01771383437776333 |
| ч 0.015009890199603998 |
| б 0.014417756916500355 |
| г 0.014004640672474556 |
| ы 0.013691852659140738 |
| з 0.012766275550501985 |
| ж 0.009459378378085954 |
| й 0.008302652894813719 |
| х 0.0070544516717929145 |
| ш 0.006824286907264256 |
| ю 0.004657393846338462 |
| э 0.002924272841639757 |
| щ 0.002479681074259612 |
| ц 0.0022986968149721195 |
| ф 0.001031806999959672 |
| 0.17087078018969903 |

Bigram frequency with whitespaces (**bigram\_frec\_with\_whitespace.xls**)



Bigram frequency without whitespaces (**bigram\_frec\_without\_whitespace.xls**)

****

**2.**

**H10**

**Graphical user interface, text

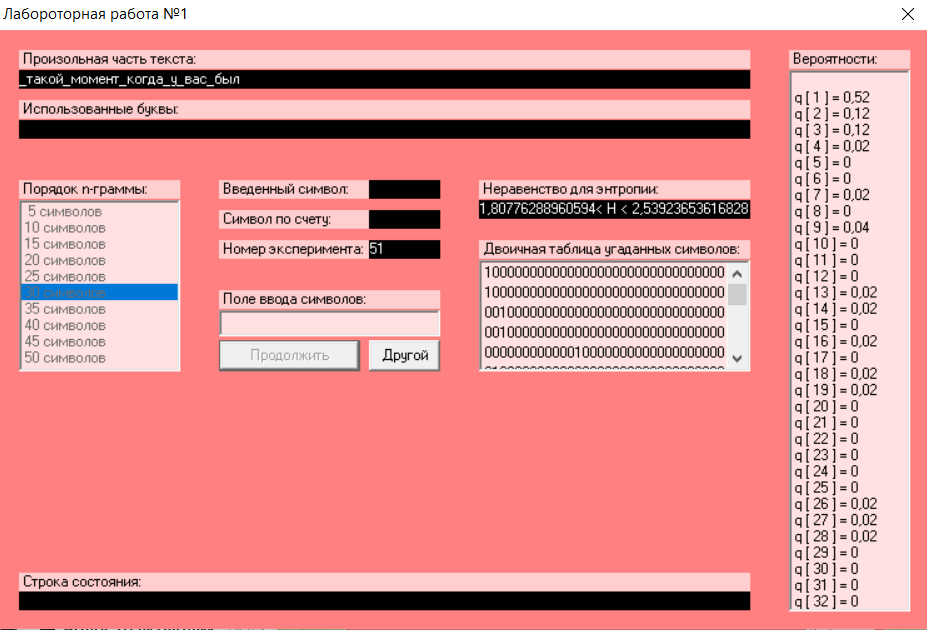
Description automatically generated**

**H20**

**Graphical user interface

Description automatically generated with medium confidence**

# H30

****

**Оцінка надлишковості мови:**

|  |  |
| --- | --- |
| **H** | **R** |
| H1 with whitespaces | 0,12618 |
| H1 without whitespaces | 0,114166 |
| H2 with whitespaces | 0,44269569 |
| H2 without whitespaces | 0,368318 |
| H10 | 0,517>R>0,427 |
| H20 | 0,475>R>0,373 |
| H30 | 0,638>R>0,492 |

**Висновки:**

Під час виконання даної лабораторної роботи я навчився вимірювати частоти символів та біграм у тексті, визначати ентропію. Також, я навчився визначати надлишковість мови.