**Міністерство освіти і науки України Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"**

**Фізико-технічний інститут**

КРИПТОГРАФІЯ

КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконав:

cтудент групи ФБ-01

Полянський Ян

**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.

***Завдання*** *:*

1.Написати програми для підрахунку частот букв і частот біграм в тексті, а також

підрахунку Н\_1 та Н\_2 за безпосереднім означенням.

2. Підрахувати частоти букв та біграм, а також значення Н\_1 та Н\_2 на добровільно.

вибраному тексті російською мового, достатньої довжини, де імовірності замінити відповідними частотам.

Також одержати значення Н\_1 та Н\_2 на тому ж тексті у якому вилучені 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.

***Хід роботи:***

*\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\*

*Маємо два (з пробілом та без) алфавіти:*

alfavit\_p = len('абвгдеёжзийклмнопрстуфхцчшщъыьэюя ')

alfavit = len('абвгдеёжзийклмнопрстуфхцчшщъыьэюя')

*Зчитуємо текст з обраного файлу, переведемо в нижній регістр:*

**

text = open('ourText.txt', mode = 'r', encoding="utf8").read()

text = text.lower()

*Використав бібілотеку* ***re****: лишаємо лиш наш алфавіт, прибираючи все, що у нього не входить, замінюючи пробілами.*

obrob\_text\_p = re.sub( r'[^а-яё ]', '',text)

obrob\_text = re.sub( r'[^а-яё]', '',text)

*«Зклеїли» раніше отримані списки за пробілом у ролі розділювача :*

obrob\_text\_p = ' '.join(obrob\_text\_p.split())

obrob\_text = ' '.join(obrob\_text.split())

*Записуємо результати у файли:*

with open("obrob\_text\_probels.txt", mode = "w") as text\_p:

    text\_p.write(obrob\_text\_p)

text\_with\_probels = obrob\_text\_p

with open("obrob\_text.txt", mode = "w") as text\_without\_probels:

    text\_without\_probels.write(obrob\_text)

text\_without\_probels = obrob\_text

**

***Частота букв з пробілом:***

print ('Частота букв з пробілом:')

for key in dictionary\_p:

    h1\_p\_letter\_name.append(key)

    h1\_p\_chastota.append(dictionary\_p[key] / dovzhyna\_p)

    print(key,":",dictionary\_p[key] / dovzhyna\_p)

*Де:*

h1\_p\_chastota = []

h1\_p\_letter\_name = []

dovzhyna\_p = len(text\_with\_probels)

dictionary\_p = dict(Counter(text\_with\_probels))

***Частота букв без пробілів:***

print ('Частота букв без пробілів:')

for key in dictionary:

    h1\_letter\_name.append(key)

    h1\_chastota.append(dictionary[key] / dovzhyna)

    print(key,":",dictionary[key] / dovzhyna)

*Де:*

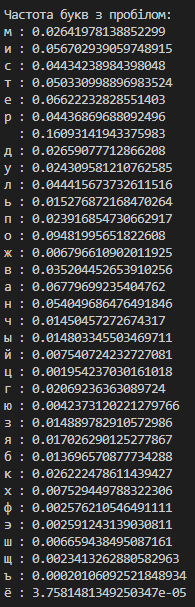
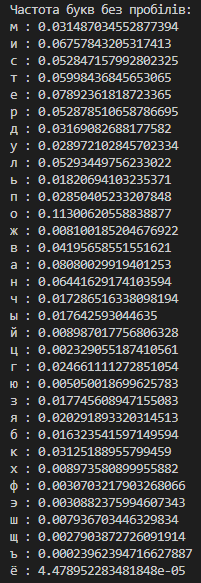
h1\_chastota = []

h1\_letter\_name = []

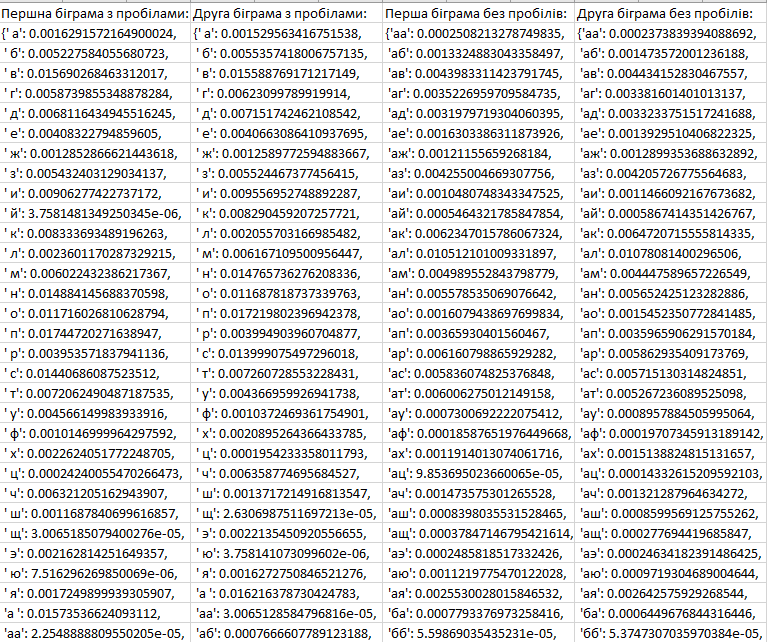
dovzhyna = len(text\_without\_probels)

dictionary = dict(Counter(text\_without\_probels))

*\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\*

*****У результаті маємо:***

***Частота біграм(фрагмент таблиці):***

**

***ТОР 15 біграм у тексті:***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Місце: | H1 з пробілами | H1 без пробілів | H2 з пробілами | H2 без пробілів | H2 з пробілами\_крк2 | H2 без пробілів\_крк2 |
| 1 | (' ', 85644) | ('о', 50461) | ('о ', 11542) | ('то', 6983) | ('о ', 5837) | ('то', 3380) |
| 2 | ('о', 50461) | ('а', 36080) | ('и ', 9952) | ('на', 5567) | ('и ', 4775) | ('на', 2792) |
| 3 | ('а', 36080) | ('е', 35242) | (' п', 9285) | ('он', 5378) | (' п', 4582) | ('но', 2747) |
| 4 | ('е', 35242) | ('и', 30176) | ('а ', 8374) | ('но', 5274) | ('а ', 4315) | ('он', 2736) |
| 5 | ('и', 30176) | ('н', 28764) | (' в', 8350) | ('по', 5084) | ('е ', 4153) | ('не', 2567) |
| 6 | ('н', 28764) | ('т', 26785) | ('е ', 8127) | ('ро', 4783) | (' в', 4148) | ('по', 2514) |
| 7 | ('т', 26785) | ('л', 23637) | (' н', 7921) | ('не', 4729) | (' н', 3929) | ('ал', 2407) |
| 8 | ('л', 23637) | ('р', 23612) | (' с', 7667) | ('ал', 4694) | (' с', 3725) | ('ст', 2336) |
| 9 | ('р', 23612) | ('с', 23598) | ('то', 6785) | ('ст', 4600) | ('то', 3378) | ('ро', 2248) |
| 10 | ('с', 23598) | ('в', 18735) | (' о', 6235) | ('ли', 4397) | (' о', 3110) | ('ос', 2186) |
| 11 | ('в', 18735) | ('д', 14151) | ('я ', 5652) | ('от', 4385) | ('я ', 2829) | ('ри', 2181) |
| 12 | ('д', 14151) | ('м', 14060) | ('на', 5555) | ('ов', 4255) | ('на', 2747) | ('ов', 2178) |
| 13 | ('м', 14060) | ('к', 13955) | ('ь ', 5224) | ('ри', 4210) | ('ь ', 2610) | ('ер', 2151) |
| 14 | ('к', 13955) | ('у', 12937) | ('по', 5084) | ('ер', 4159) | (' и', 2543) | ('от', 2146) |
| 15 | ('у', 12937) | ('п', 12728) | ('но', 5021) | ('ко', 4135) | ('по', 2523) | ('ли', 2087) |

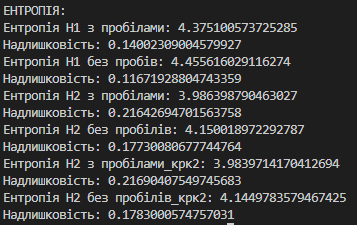
***Ентропія та надлишковість:***

***H:***

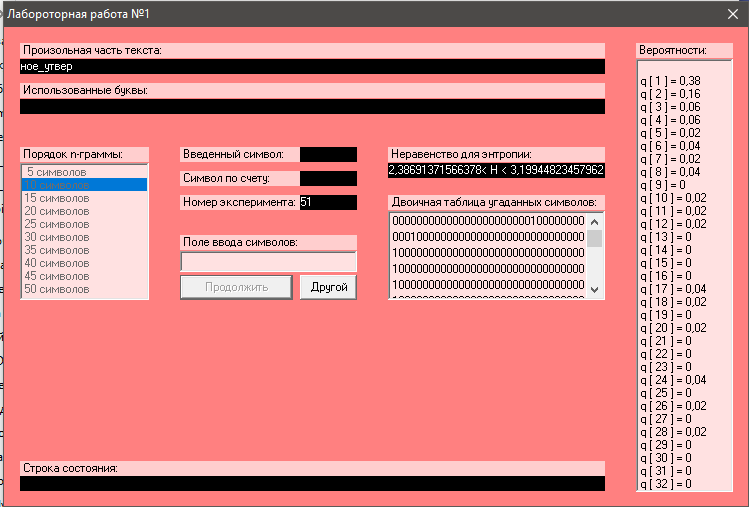
******

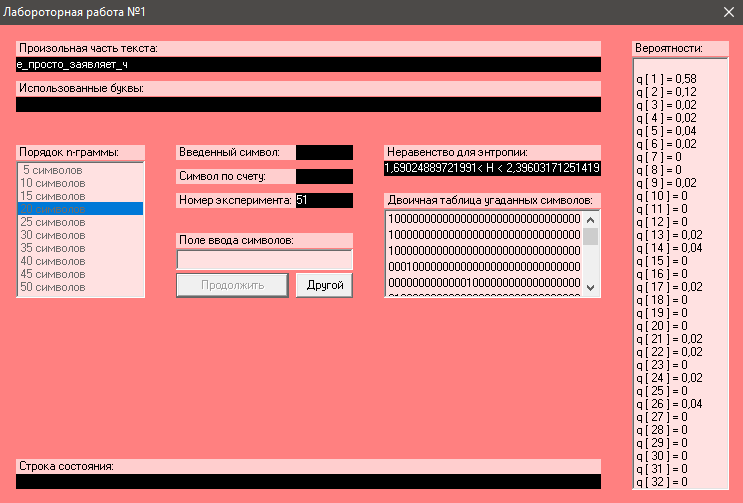
***R:***

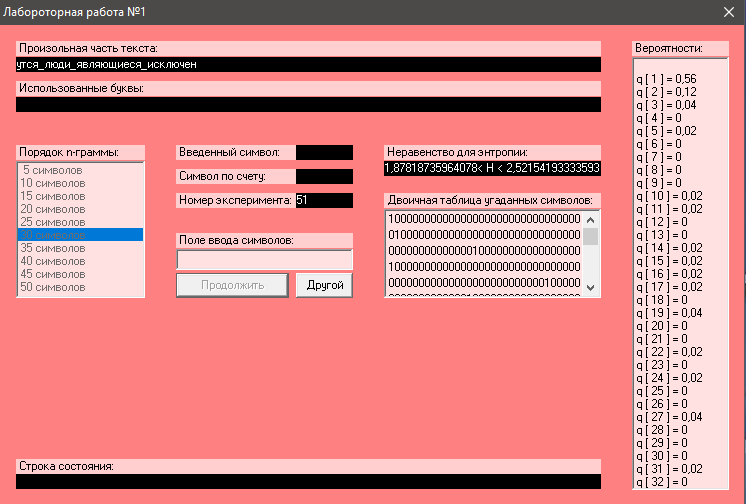
******

**

***CoolPinkProgram:***

**

**

**