Міністерство освіти і науки України
Національний технічний університет України
“Київський політехнічний інститут ім. Ігоря Сікорського”
Навчально-Науковий Фізико-технічний інститут

КРИПТОГРАФІЯ

КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1
Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконав:
студент 3го курсу
групи ФБ-03
Заболотний Максим Олександрович

Тема: експериментальна оцінка ентропії на символ джерела відкритого тексту

Мета: 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.

Хід роботи

Частоти букв в тексті з пробілами:

(' ', 0.15858840767300678)
('о', 0.09476878616901291)
('е', 0.07925655983008074)
('а', 0.06356624838384145)
('и', 0.058099180374152466)
('н', 0.05593030646564767)
('т', 0.052897068255822875)
('с', 0.05005928931011567)
('л', 0.039679678462271294)
('в', 0.0363568709722723)
('р', 0.03414600951814224)
('м', 0.029997929579646754)
('д', 0.02719055540835784)
('к', 0.02316699026033726)
('п', 0.02086201879015059)
('у', 0.020209040063357758)
('ь', 0.018710519038456247)
('б', 0.018099527857288675)
('я', 0.016842797181332626)
('з', 0.015171982434727465)
('ы', 0.014992449481019458)
('г', 0.014000674696422806)
('ч', 0.012448583354689067)
('ж', 0.00912867155749181)
('й', 0.00873485733645489)
('х', 0.006940975645775691)
('ш', 0.006506621725514384)
('ю', 0.004287073192979103)
('э', 0.004116227317676322)
('щ', 0.0026017799823652306)
('ц', 0.0018865438603349449)
('ф', 0.0007557758212546747)

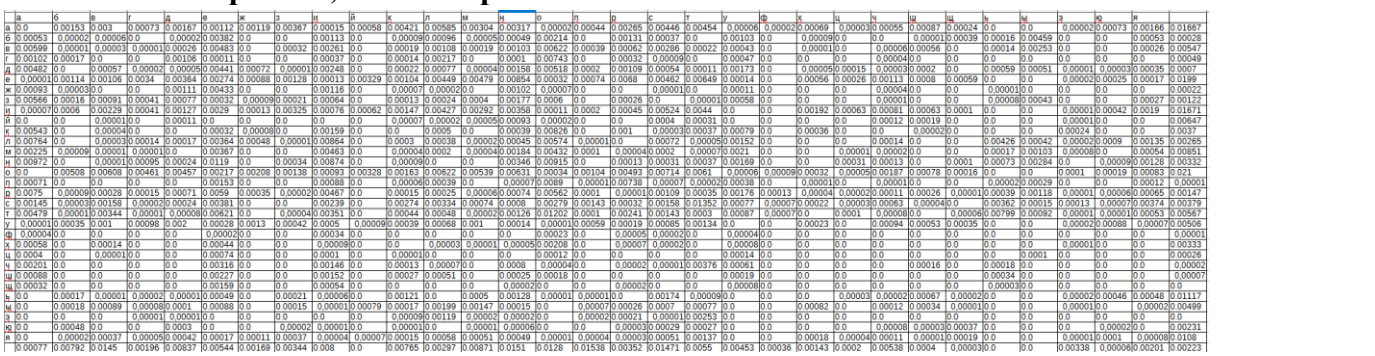
Частоти букв в тексі без пробілів:

('о', 0.11263071133465372)
('е', 0.09419475623207209)
('а', 0.07554715072090194)
('и', 0.0690496552507369)
('н', 0.06647199417703266)
('т', 0.06286705429091091)
('с', 0.05949441363372778)
('л', 0.04715846420957176)
('в', 0.04320937731761499)
('р', 0.040581814928064674)
('м', 0.03565190906947812)
('д', 0.032315403847907674)
('к', 0.027533481201830173)
('п', 0.024794071035383475)
('у', 0.024018019537225523)
('ь', 0.02223705878202933)
('б', 0.021510908599717455)
('я', 0.0200173105943935)
('з', 0.01803158237072548)
('ы', 0.01781821122710777)
('г', 0.016639507732122853)
('ч', 0.014794879780847186)
('ж', 0.01084923435894877)
('й', 0.010381194431013152)

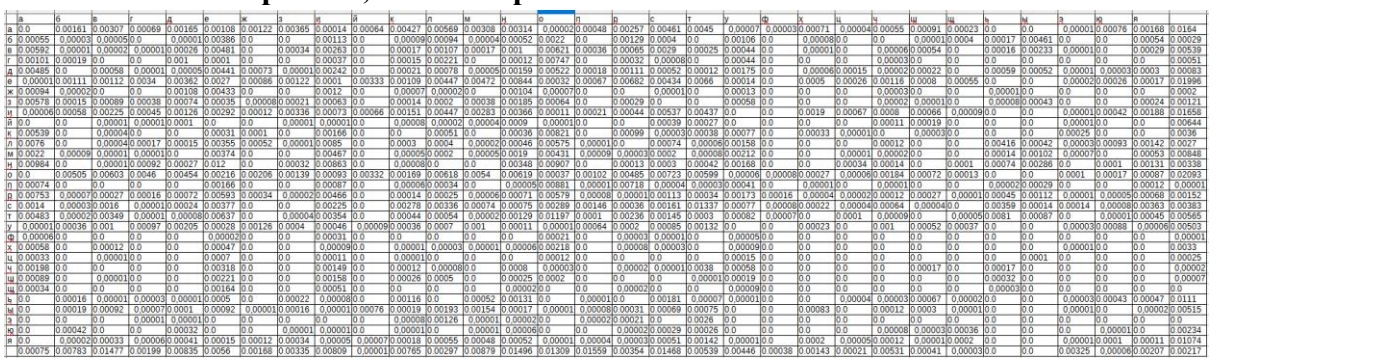
('х', 0.00824920372986525)
('ш', 0.00773298322111273)
('ю', 0.00509509642138736)
('э', 0.0048920496879447703)
('щ', 0.0030921608474275874)
('ц', 0.0022421177430151065)
('ф', 0.0008982236852293825)

Частоти біграм:

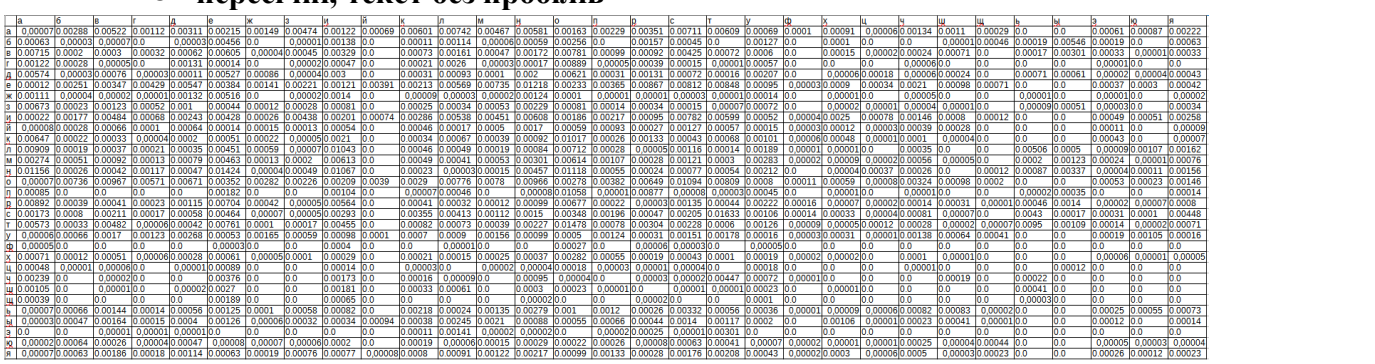
- **пересічні, текст із пробілами**



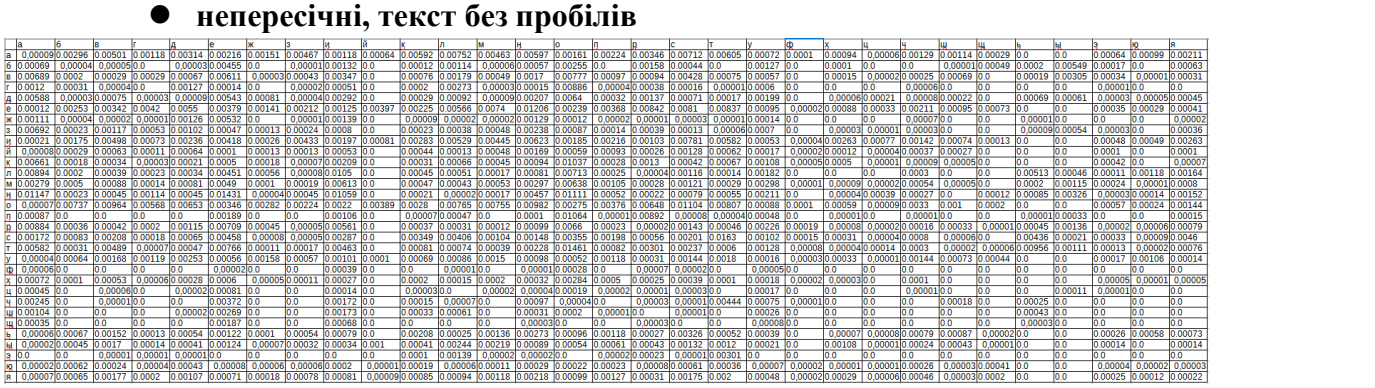
- **непересічні, текст із пробілами**



- **пересічні, текст без пробілів**



- **непересічні, текст без пробілів**

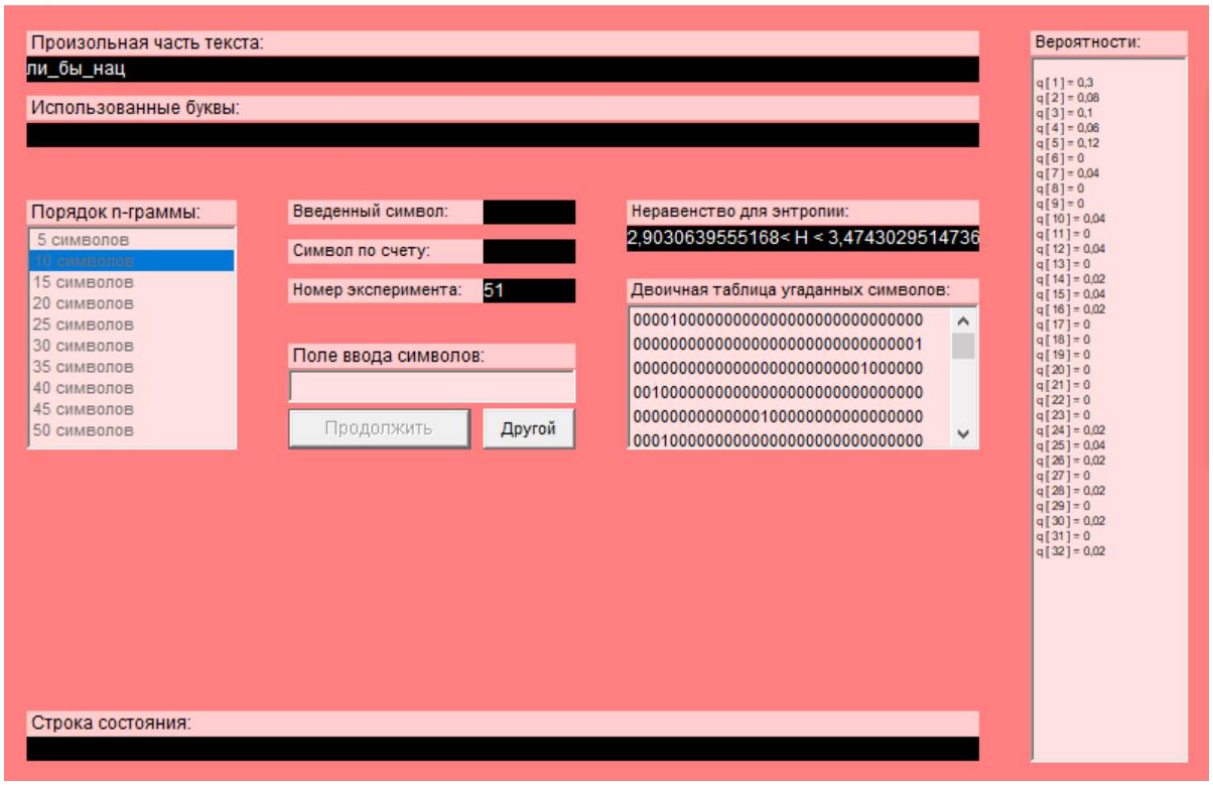


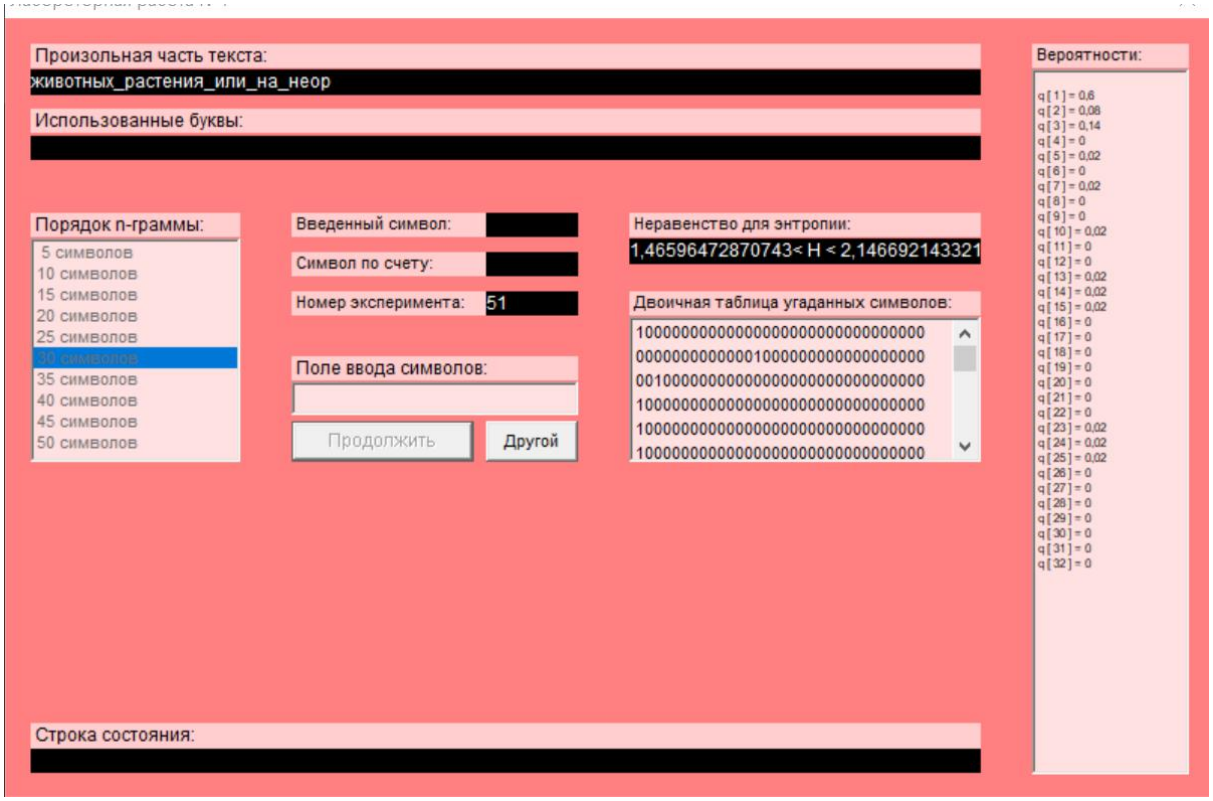
Значення ентропій та надлишковість:

```
H1 для тексту з пробілами: 4.363132070543333
H2 для тексту з пробілами(з перетином): 3.9463793212258014
H2 для тексту з пробілами(без перетину): 3.946656736379984
H1 для тексту без пробілів: 4.435653426951057
H2 для тексту без пробілів(з перетином): 4.096810973121039
H2 для тексту без пробілів(без перетину): 4.0964554962740385

Надлишковість російської мови: 0.10466740737516811
```

КулПінкПрограм ентропії:





Висновки

В ході виконання даної лабораторної роботи я засвоїв поняття ентропії на символ джерела та його надлишковості, вивчив та порівняв різні моделі джерела відкритого тексту для наближеного визначення ентропії, погрався з кулпінк програм та побачив, що зі збільшенням довжини тексту - значення ентропії зменшується, опанував

деякі методи фільтрації тексту, що зможе вподальшому полегшити виконання наступних лаб робіт.