

Міністерство освіти і науки, молоді та спорту України

Національний технічний університет України

“Київський політехнічний інститут”

Фізико-Технічний інститут

**КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1**

**за семестровий курс предмету  
«Симетрична криптографія»**

**Роботу виконали:**

Студенти групи ФІ-03

Піжук Богдан

Швець Катерина

**Приймав:**

Чорний Олег Миколайович

Київ-2023

КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела

відкритого тексту

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:

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 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення на тому ж тексті, в якому вилучено всі пробіли.
3. За допомогою програми CoolPinkProgram оцінити значення .
4. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи

* Очистка вхідного текстового файлу та редагування;
* Написання функцій для підрахунку частот літер та частот біграм. Визначили частоти та значення ентропії;
* Результати оформили у вигляді таблиць (див. нижче);
* Було проведено по 50 експерименів для умовних ентропій джерел;
* Обрахували оцінки надлишковості російської мови для даних умовних ентропій джерел;
* Оформлення звіту.

Таблиці частот букв і біграм тексту, одержані значення H1 та H2

|  | Текст з пробілами | | | Текст без пробілів | | |
| --- | --- | --- | --- | --- | --- | --- |
| Літери | Біграми з | Біграми без | Літери | Біграми з | Біграми без |
| H | 4.356764105666061 | 3.970778960772621 | 3.970778960772621 | 4.513118781529859 | 4.1919475573122496 | 4.191943806270274 |
| R | 0.12864717886678778 | 0.20584420784547575 | 0.20584420784547575 | 0.08903109631167849 | 0.15385921455645768 | 0.1538599717008542 |

| Текст з пробілами | | | | Текст без пробілів | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Літера | Частота | Літера | Частота | Літера | Частота | Літера | Частота |
| \_ | 0.17192241820532642 | я | 0.018241818200896088 | о | 0.11266149014507101 | ь | 0.019636676946528837 |
| 0 | 0.09512700427991656 | ь | 0.0165804504230367 | е | 0.08131930190829866 | г | 0.01958834985774576 |
| е | 0.06866287291877253 | г | 0.016539644898668226 | а | 0.08116672638514066 | ы | 0.018908318679869626 |
| а | 0.06853404404898061 | ы | 0.015965452877197524 | и | 0.06566891939939094 | б | 0.017542733228256436 |
| и | 0.055448295320072706 | б | 0.014812405345756867 | н | 0.06384353621735592 | з | 0.01687927133796307 |
| н | 0.05390701237106912 | з | 0.014252203789783929 | т | 0.057043914825577184 | ч | 0.014047303935274839 |
| т | 0.04816567509242451 | ч | 0.011861000061791223 | с | 0.05230095625500962 | й | 0.011366531281779375 |
| с | 0.04416090434368978 | й | 0.009597459331465606 | л | 0.048450668052963726 | ж | 0.010043059436105721 |
| л | 0.040909869923647035 | ж | 0.008479970899831765 | в | 0.04498699656118244 | ш | 0.009176623772923437 |
| в | 0.0379852796985521 | ш | 0.007748386141511227 | р | 0.0429047894216145 | х | 0.008689210563196988 |
| р | 0.036227144534333185 | х | 0.007336833281452024 | к | 0.03320070999397292 | ю | 0.006207269360694751 |
| к | 0.02803339524114316 | ю | 0.005241178137099567 | м | 0.03044192361144192 | ц | 0.003578275730895439 |
| м | 0.025703982735765577 | ц | 0.0030213576114544605 | д | 0.02988685247741917 | э | 0.0030867201992732986 |
| д | 0.025235302141590508 | э | 0.0026063071350208166 | у | 0.026790466860389255 | щ | 0.002939667771976226 |
| у | 0.022620833901695996 | щ | 0.0024821417537281678 | п | 0.024891902658197 | ф | 0.001861283305131025 |
| п | 0.021017759730077287 | ф | 0.001571595624248741 | я | 0.02160427984698263 |  |  |

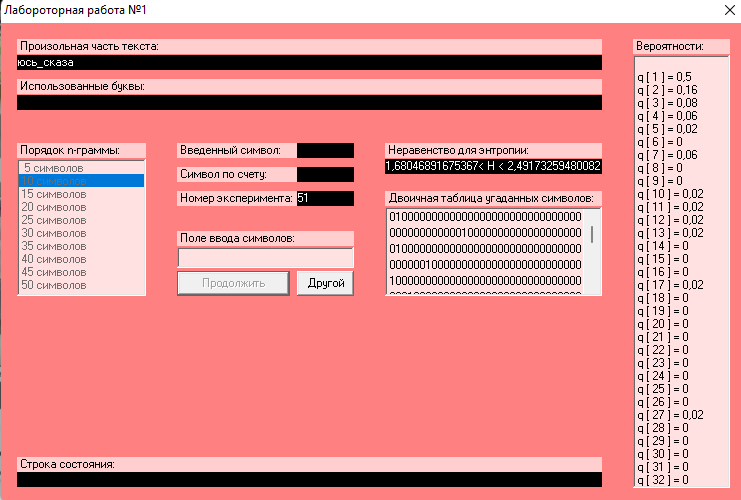
-Біграми (дуже багато, в таблиці наведено 5 найчастіших)

| Текст з пробілами | | | | Текст без пробілів | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| з перетином | | без перетину | | з перетином | | без перетину | |
| 'о\_' | 0.021034094201356726 | 'о\_' | 0.02107896801663 | 'то' | 0.01630971333731917 | 'то' | 0.016400844481940156 |
| 'и\_' | 0.018540875209055566 | 'и\_' | 0.018507054109291186 | 'ст' | 0.013205731320531709 | 'ст' | 0.013100792426725728 |
| 'а\_' | 0.016745431090213488 | 'а\_' | 0.01660784841796982 | 'на' | 0.01226887553832962 | 'на' | 0.012265423601033372 |
| 'е\_' | 0.016596782307647016 | 'е\_' | 0.016477270739990696 | 'ов' | 0.011162184441152063 | 'ов' | 0.011228461637240051 |
| '\_с' | 0.016061646690407723 | '\_с' | 0.01585119740896579 | 'го' | 0.010625062997855657 | 'но' | 0.010572593550952665 |

Оцінки для

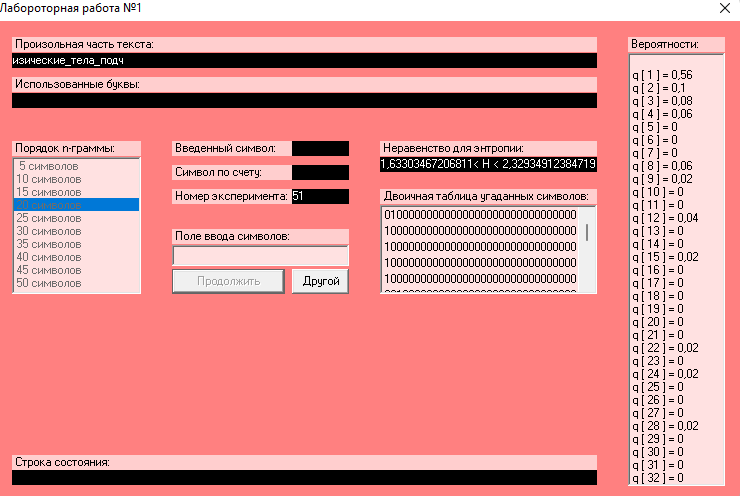
1,68046891675367<<2,49173259480082

R=



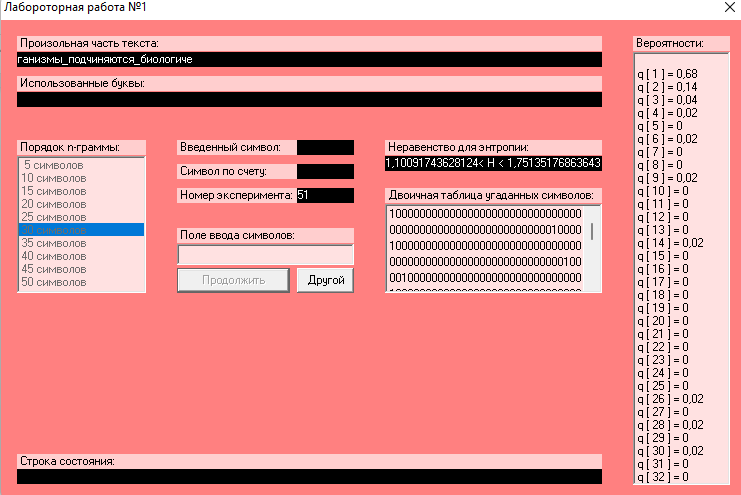
1,63303467206811<<2,32934912384719

R=



1,10091743628124<

R=



Висновки: відомо, що пробіл є найвживанішим символом в російській мові. В даному комп’ютерному практикумі ми розглянули над одним текстом два випадки (з пробілом та без пробілу) та порівняли оцінки ентропії та надлишковості в літерах та біграмах(з перетином та без). Переконалися на власному дослідженні, що найчастіший символ є “ “, а найчастіші літери (“о”, “е”, “а”). Навчилися оцінювати ентропію та надлишковість на символ джерала.