**КРИПТОГРАФІЯ**

**Комп’ютерний практикум №1**   
Експериментальна оцінка ентропії на символ джерелавідкритого тексту

Виконав: ст. ФБ-12 Слєпий Роман

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:**0. Уважно прочитати методичні вказівки до виконання комп’ютерного практикуму.  
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**:

1.Ознайомився з методичними вказівками до виконання комп’ютерного практикуму та рекомендаціями стосовно виконання(лайфхаками)  
2. Створив програми(4 файли формату .py) для обрахунку частоти та ентропії символів і біграм(з повторами і без) у тексті з пробілами та без.

Для підрахунку ентропії Н(1) символів користався формулою

Де n – кількість літер алфавіту, p(i) – імовірність (частота) появи літери в тексті

Далі обчислюю H(2). Для цього треба порахувати частоту біграм. Роблю це аналогічно до частоти символів, але з урахуванням того, що частота біграм – відношення кількості появ деякої біграми до загальної кількості біграм у тексті.

Для підрахунку ентропії Н(2) біграм(з повторами і без) користався формулою

Де p(i,j) – частота появи деякої біграми в тексті.

Де 𝐻(𝑥1, 𝑥2, 𝑥3 … , 𝑥𝑛) в свою чергу, – ентропія n-грами відкритого тексту  
(𝑥1, 𝑥2, 𝑥3 … , 𝑥𝑛).  
Усі попередні розрахунки були реалізовані у коді(усі 4 фрагменти)

Таблиці з результатами виконання коду:

Таблиця для символів(літер) з пробілом: Таблиця для символів(літер) без пробілу:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Кількість | Частота |  |  | Кількість | Частота |
| «\_» | 137530 | 0.16217848 |  | А | 54592 | 0.07580200 |
| А | 54592 | 0.06437986 |  | Б | 13679 | 0.01899355 |
| Б | 13679 | 0.01613152 |  | В | 29428 | 0.04086133 |
| В | 29428 | 0.03470418 |  | Г | 14105 | 0.01958506 |
| Г | 14105 | 0.01663390 |  | Д | 24446 | 0.03394373 |
| Д | 24446 | 0.02882895 |  | Е | 56302 | 0.07817638 |
| Е | 56302 | 0.06639645 |  | Ё | 5 | 0.00000694 |
| Ё | 5 | 5.896455876231032e-06 |  | Ж | 6466 | 0.00897816 |
| Ж | 6466 | 0.00762529 |  | З | 13069 | 0.01814655 |
| З | 13069 | 0.01541215 |  | И | 49147 | 0.06824152 |
| И | 49147 | 0.05795862 |  | Й | 7419 | 0.01030142 |
| Й | 7419 | 0.00874916 |  | К | 22018 | 0.03057240 |
| К | 22018 | 0.02596563 |  | Л | 38145 | 0.05296504 |
| Л | 38145 | 0.04498406 |  | М | 22475 | 0.03120696 |
| М | 22475 | 0.02650456 |  | Н | 47645 | 0.06615597 |
| Н | 47645 | 0.05618732 |  | О | 80993 | 0.11246029 |
| О | 80993 | 0.09551433 |  | П | 19952 | 0.02770372 |
| П | 19952 | 0.02352921 |  | Р | 34322 | 0.04765674 |
| Р | 34322 | 0.04047563 |  | С | 37102 | 0.05151682 |
| С | 37102 | 0.04375406 |  | Т | 39123 | 0.05432301 |
| Т | 39123 | 0.04613740 |  | У | 20185 | 0.02802725 |
| У | 20185 | 0.02380399 |  | Ф | 2362 | 0.00327968 |
| Ф | 2362 | 0.00278548 |  | Х | 7419 | 0.01030142 |
| Х | 7419 | 0.00874916 |  | Ц | 2171 | 0.00301447 |
| Ц | 2171 | 0.00256024 |  | Ч | 9415 | 0.01307290 |
| Ч | 9415 | 0.01110302 |  | Ш | 5693 | 0.00790484 |
| Ш | 5693 | 0.00671370 |  | Щ | 2092 | 0.00290478 |
| Щ | 2092 | 0.00246707 |  | Ъ | 188 | 0.00026104 |
| Ъ | 188 | 0.00022170 |  | Ы | 14363 | 0.01994329 |
| Ы | 14363 | 0.01693815 |  | Ь | 15212 | 0.02112215 |
| Ь | 15212 | 0.01793937 |  | Э | 3255 | 0.00451963 |
| Э | 3255 | 0.00383859 |  | Ю | 3433 | 0.00476678 |
| Ю | 3433 | 0.00404850 |  | Я | 14224 | 0.01975029 |
| Я | 14224 | 0.01677423 |  | - | - | - |

Для біграм без повторів,з пробілами: Для біграм без повторів,без пробілів:

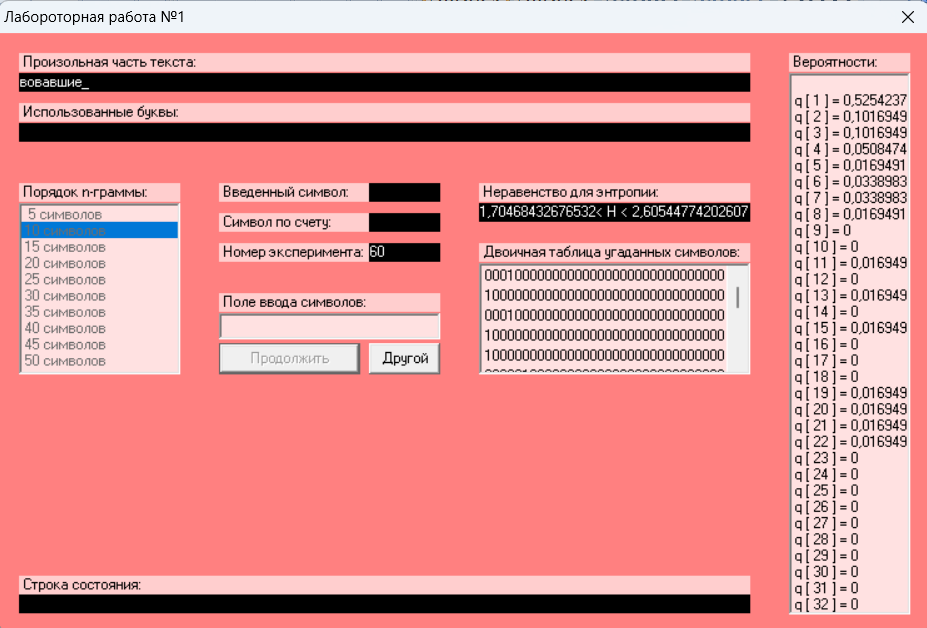
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Кількість | Частота |  |  | Кількість | Частота |
| о\_ | 18602 | 0.02193720 |  | то | 9316 | 0.012935457 |
| и\_ | 17113 | 0.02018123 |  | но | 8541 | 0.011859354 |
| е\_ | 14355 | 0.01692874 |  | на | 8352 | 0.011596924 |
| \_п | 13947 | 0.01644759 |  | не | 8155 | 0.011323385 |
| \_н | 13882 | 0.01637094 |  | ст | 8125 | 0.011281729 |
| \_с | 12943 | 0.01526358 |  | по | 8030 | 0.01114982 |
| а\_ | 12878 | 0.01518693 |  | ро | 7975 | 0.011073451 |
| \_в | 12618 | 0.01488031 |  | ал | 7959 | 0.011051235 |
| \_о | 9574 | 0.01129055 |  | ли | 7653 | 0.010626348 |
| \_и | 9170 | 0.01129055 |  | он | 7131 | 0.00990154 |
| То | 9032 | 0.01065137 |  | ра | 6850 | 0.0095113657 |
| ь\_ | 8695 | 0.01025395 |  | ни | 6819 | 0.0094683216 |
| я\_ | 8431 | 0.00994262 |  | ос | 6608 | 0.0091753438 |
| но | 8341 | 0.00983648 |  | ко | 6508 | 0.0090364917 |
| на | 8277 | 0.00976100 |  | ов | 6410 | 0.0089004167 |
| не | 8097 | 0.00954873 |  | го | 6399 | 0.008885143 |
| по | 8027 | 0.00946618 |  | от | 6099 | 0.0084685868 |
| ст. | 7911 | 0.00932938 |  | од | 6078 | 0.0084394279 |
| ро | 7902 | 0.00931877 |  | ор | 6024 | 0.0083644478 |
| ал | 7712 | 0.00909470 |  | ер | 5913 | 0.008210322 |
| ли | 7212 | 0.00850506 |  | ло | 5507 | 0.00764658 |
| м\_ | 7016 | 0.00827392 |  | ен | 5497 | 0.00763270 |
| ра | 6824 | 0.00804749 |  | ил | 5357 | 0.00743830 |
| л\_ | 6729 | 0.00793546 |  | ол | 5336 | 0.00740915 |
| \_т | 6668 | 0.00786352 |  | ре | 5229 | 0.00726057 |
| ни | 6568 | 0.00774559 |  | ом | 5208 | 0.00723141 |
| \_к | 6341 | 0.00747789 |  | во | 5192 | 0.00720920 |
| го | 6335 | 0.00747082 |  | пр | 5142 | 0.00713977 |
| ко | 6323 | 0.00745667 |  | ка | 5096 | 0.00707590 |
| й\_ | 5850 | 0.00689886 |  | ес | 4927 | 0.00684124 |

Для біграм з повторами,з пробілами: Для біграм з повторами,без пробілів:

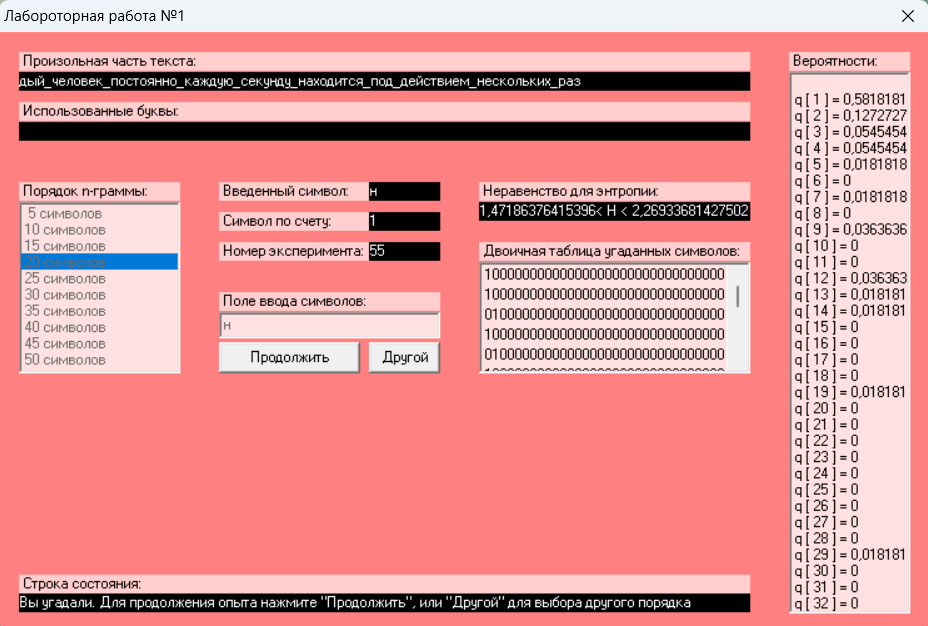
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Біграма | Частота |  | Біграма | Частота |
| о\_ | 0.02193717 |  | то | 0.01293544 |
| и\_ | 0.02018121 |  | но | 0.01185934 |
| е\_ | 0.01692872 |  | на | 0.01159691 |
| \_п | 0.01644757 |  | не | 0.01132337 |
| \_н | 0.01637092 |  | ст | 0.01128171 |
| \_с | 0.01526357 |  | по | 0.01114980 |
| а\_ | 0.01518691 |  | ро | 0.01107344 |
| \_в | 0.01488030 |  | ал | 0.01105122 |
| \_о | 0.01129053 |  | ли | 0.01062633 |
| \_и | 0.01081410 |  | он | 0.00990153 |
| то | 0.01065136 |  | ра | 0.00951135 |
| ь | 0.01025394 |  | ни | 0.00946831 |
| я | 0.00994260 |  | ос | 0.00917533 |
| но | 0.00983647 |  | ко | 0.00903648 |
| на | 0.00976099 |  | ов | 0.00890040 |
| не | 0.00954872 |  | го | 0.00888513 |
| по | 0.00946617 |  | от | 0.00846858 |
| ст | 0.00932937 |  | од | 0.00843942 |
| ро | 0.00931876 |  | ор | 0.00836444 |
| ал | 0.00909469 |  | ер | 0.00821031 |
| ли | 0.00850505 |  | ло | 0.00764657 |
| м\_ | 0.00827391 |  | ен | 0.00763269 |
| ра | 0.00804748 |  | ил | 0.00743829 |
| л\_ | 0.00793545 |  | ол | 0.00740914 |
| \_т | 0.00786351 |  | ре | 0.00726056 |
| ни | 0.00774558 |  | ом | 0.00723140 |
| \_к | 0.00747789 |  | во | 0.00720919 |
| го | 0.00747081 |  | пр | 0.00713976 |
| ко | 0.00745666 |  | ка | 0.00707589 |
| й\_ | 0.00689885 |  | ес | 0.00684123 |

3. За допомогою CoolPinkProgram оцінив значення H (10) , H (20) , H (30) та прорахував надлишковість.

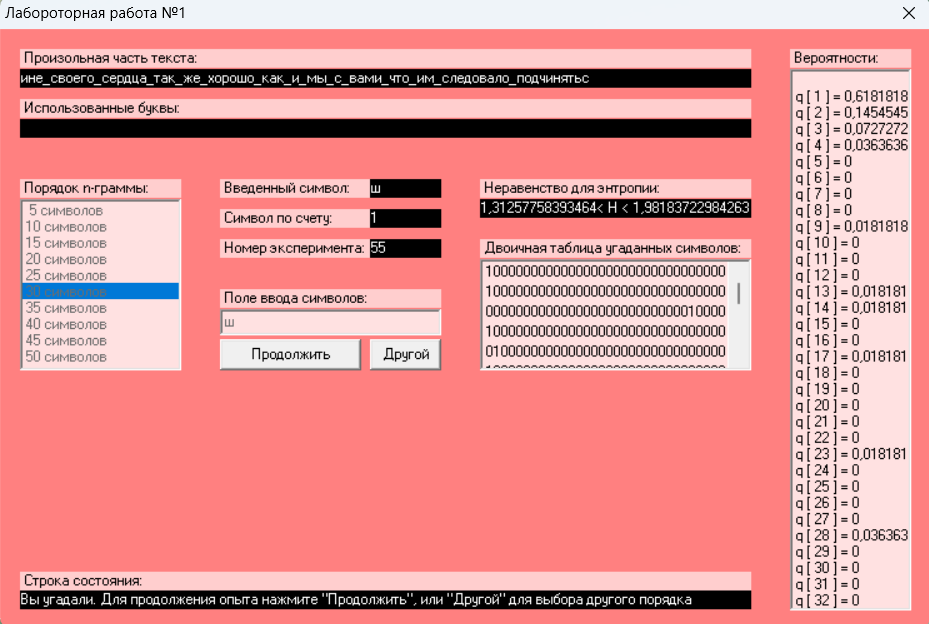
3.1 Для H(10)



3.2 Для Н(20)



3.3 Для Н(30)



Користуючись формулою розрахуємо надлишковість для даного тексту.

- для словника, до якого входять і пробіли

*–* для словника, до якого не входять пробіли

|  |  |  |
| --- | --- | --- |
|  | Ентропія | Надлишковість |
| H1(з пробілами) | 4.38958257 | 0,137176 |
| H1(без пробілів) | 4.51882074 | 0,104189 |
| H2(з пробілами, без повторів) | 3.98739756 | 0,216230 |
| H2(з пробілами з повторами) | 3.98739371 | 0,216231 |
| H2( без пробілів, без повторів) | 4.18509826 | 0,170346 |
| H2(без пробілів, з повторами) | 4.18509345 | 0,170358 |

Для CoolPinkProgram:

|  |  |  |
| --- | --- | --- |
|  | Ентропія | Надлишковість |
| H(10) | 1,70468432<Н<2,60544774 | 0,4878690<R<,664924 |
| H(20) | 1,47186376<Н<2,26933681 | 0,5539350<R<,710688 |
| H(30) | 1,31257758<Н<1,98183722 | 0,6104470<R<,741997 |

4.  
**Висновки**:

*Під час виконання комп'ютерного практикуму я навчився експериментально визначати значення частоти літер та біграм(n-грам) в тексті. З використанням отриманих даних зміг розрахувати ентропію та надлишковість для ВТ. А за допомогою CoolPinkProgram.exe розрахувати ентропію(експериментально) та надлишковість до наданого тексту.*