«Криптографія» Комп 'ютерний практикум №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконали: студенти групи ФБ-01 Корабельський Тарас та Оліферчук Владислав варіант №10

Мета роботи

Засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела івдкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо, оцінки ентропії, на символ джерела.

Постановка задачі:

Визначити надлишковість російської мови в різних моделях джерела, вираховуючи за допомогою самостійно написанної програми частоту входження буквенних символів в текст(розмір від 1 мб, визначення ентропію H1 та H2 для букв та біграм(з букв, які перетинаюся, і не перетинаються) з урахуванням пробілів і без них, за допомогою спеціальної програми CoolPinkProgram оцінити H^{10}, H^{20}, H^{30}

Для лабораторної роботи використовувався текст: «Война и Мир» Льва Толстого

Джерело	Н	R
Н1 текст з пробілами	4.254	0.1492
Н1 текст без пробілів	4.322	0.1356
H2 перехресна біграма з пробілами	3.7879	0.24242
H2 не перехресна біграма з пробілами	3.7883	0.2432
H2 перехресна біграма без пробілів	3.924	0.21572
H2 не перехресна біграма без пробілів	3.9239	0.21576

Частоти літер у тексті з пробілами:

- { « » 0.16300261133460373
- a 0.06892681205448514
- б 0.014898440256898864
- в 0.03793746912273273
- г 0.0164946008892653
- д 0.025448232055896677
- e 0.06898779024631237
- ж 0.008715364528195355
- 3 0.01423502011433411
- и 0.055463617757075305
- й 0.00969948479074035

- 0.028951090408638577 K
- 0.04248711976850872 Л
- 0.025002752487825533 Μ
- 0.05502773660808808 Н
- 0.09576455642600043 0
- 0.02172348083845014 П
- 0.03723452607805773 р
- 0.04402004375749877 С
- 0.04858098666102054 Т
- У 0.023269955536735127
- 0.001648669630884325 ф
- 0.007236078763497777 Χ
- 0.002987366786646905
- Ц Ч 0.012180393817488885
- 0.007745924200719881 Ш
- 0.0023837956101348014 Щ
- 0.016190274543016445 Ы
- 0.01698637871409415 Ь
- 0.0026694897311031122 Э
- 0.005325993365798574 Ю
- 0.018773943115251605} Я
- H1 = 4.225

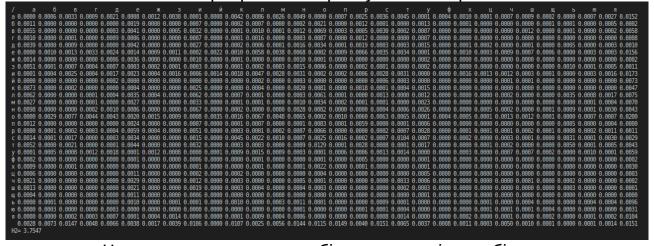
Частоти літер у тексті без пробілів:

- 0.08235009211246153 {a
- б 0.01779986468136374
- 0.045325672022973135 В
- 0.019706872581246832 Γ
- 0.030404195282465848 Д
- 0.08242294561553452 е
- 0.010412654383655991 Ж
- 0.0170072455507084 3
- 0.06626498303120491 И
- й 0.011588428974917483
- 0.03458922429232603 Κ
- 0.050761352835586505 Л
- 0.0298719599683492 Μ
- 0.06574421539812753 Н
- 0 0.11441440286772975
- 0.02595406046975669 П
- 0.04448583302921493 р
- 0.05259280895450501 С
- Т 0.05804198115657634
- 0.02780170625602475 ٧

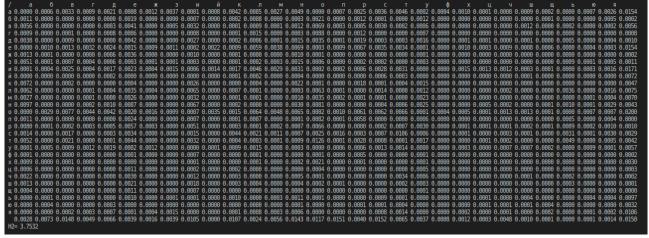
- ф 0.001969742860862626
- x 0.008645282364662813
- ц 0.0035691470811041624
- ч 0.014552487238831997
- ш 0.009254418598689852
- щ 0.0028480323145760296
- ы 0.019343279635354725
- ь 0.02029442259214113
- э 0.0031893644678624984
- ю 0.006363213837848339
- я 0.022430109543336706}

H1 = 4.321

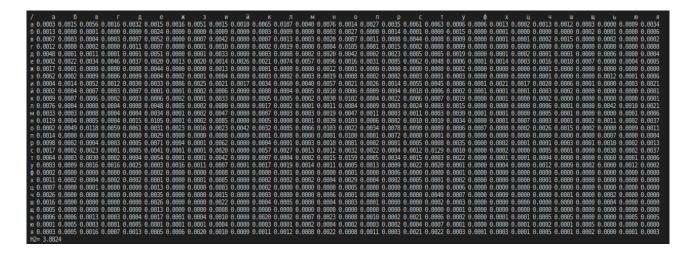
Частоти перехресних біграм у тексті з пробілами



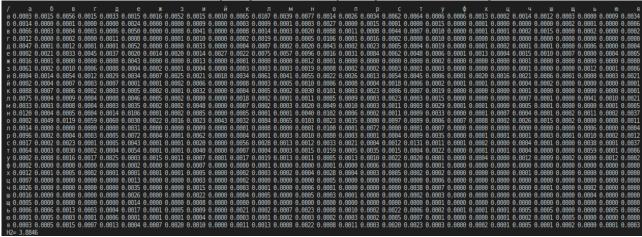
Частоти не перехресних біграм у тексті з пробілами



Частоти перехресних біграм у тексті без пробілів



Частоти не перехресних біграм у тексті без пробілів



Найбільш часті біграми для різних джерел: Перехресні з пробілами:

o «» - 0.0217

и «» - 0.0187

e «» - 0.0169

a «» - 0.0162

«» c - 0.0155

«» п - 0.0152

«» H - 0.0151

«» 0 - 0.0120

«» и - 0.0113

Не перехресні з пробілами:

 $\ll B - 0.0176$

 $\ll c - 0.0165$

 $p \ll -0.0158$

 $\Pi \ll -0.0155$

 $B \ll -0.0152$

 $H \ll > -0.0143$

 $\ll 0 - 0.0118$

«» и -0.0112

Перехресні без пробілів:

TO - 0.0159

CT - 0.0129

Ha - 0.0119

08 - 0.0118

ла - 0.0107

He - 0.0105

or - 0.0105

HO - 0.0103

KO - 0.0102

Не перехресні без пробілів:

TO - 0.0159

CT - 0.0131

Ha - 0.0120

BO - 0.0119

ла - 0.0107

ro - 0.0106

He - 0.0106

HO - 0.0103

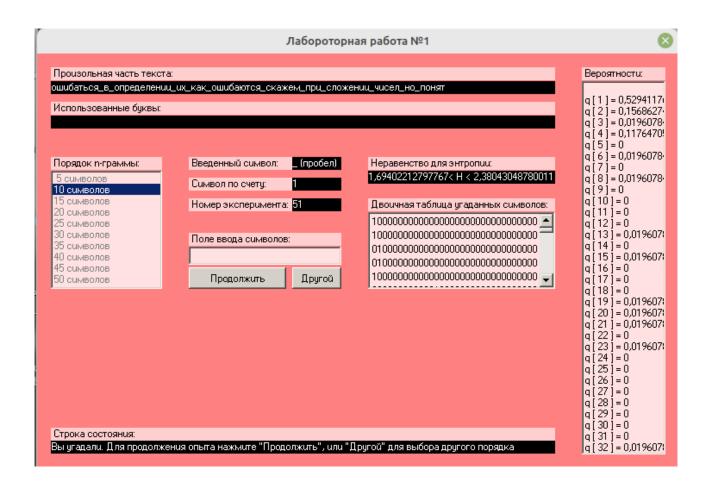
OH - 0.0102

Завдання 2

H(10)

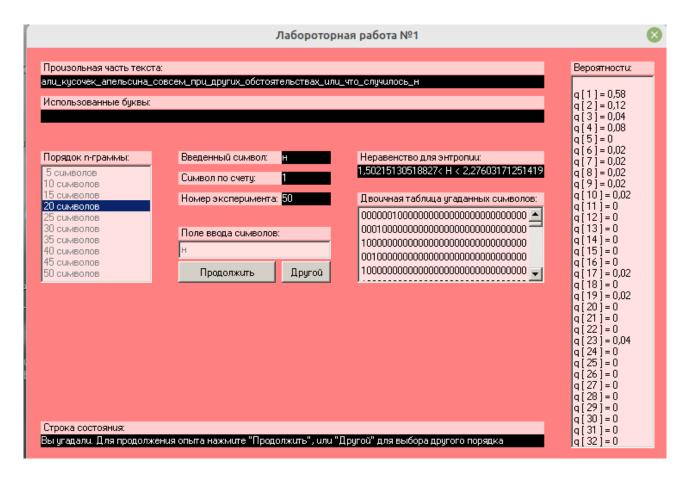
1.69402212797767<H<2.38043048780011

0.661195574404466>R>0.523913902439978



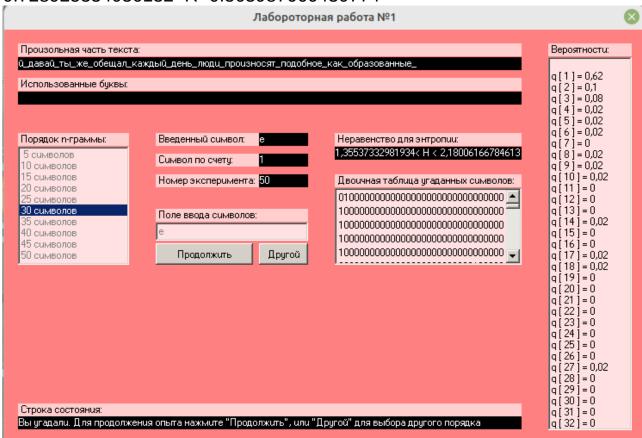
H(20)

- 1.50215130518827<H<2.27603171251419
- 0.699569738962346>R>0.5447936457497162



H(30)

- 1.35537332981934<H<2.18006166784613
- 0.728925334036132>R>0.563987666430774



Висновок:

Впродовж виконання лабораторної роботи ознайомились з поняттям ентропії на символ джерела та навчились її визначати, освоїли вимірювання частоти повторювання символів в тексті. Надлишковість може існувати в диапазоні 0.11-0.24. Також помітили за допомогою CoolPinkProgram що H⁽ⁿ⁾ менше, чим більше порядок н-грамми.