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:
Іванілов Ігор
Березовський Максим
Варіант 12
Група:
ФБ-06

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.

Хід роботи

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.

Був обран збірник Лавкрафта як текст для прикладу, та перероблен у txt файл. Відкривається файл за допомогою моєї функції open_text(). Нижче наведені результати підрахунків, всі результати виписані в окремі файли.

Частота літер з пробілами

```
` : 0.15671903909133184
г : 0.01398150804879738
о : 0.09826137071959316
в : 0.03722506950559508
а : 0.05979831009884141
р : 0.035481245856872894
д : 0.024355403627153188
л : 0.039362181041050354
к : 0.02683633287490972
ф : 0.0016152011951994143
т : 0.04972865609324138
у : 0.020469521425532548
с : 0.046535356333666435
е : 0.07428936094428669
ь : 0.014862077153684044
п : 0.023432784873999465
б : 0.013650057880104085
н : 0.058494771002562555
и : 0.060226227107676784
з : 0.014891759258343145
```

Частота літер без пробілів

```
г : 0.016579892938329545
о : 0.11652269560753832
в : 0.04414313998680062
а : 0.07091149079709613
р : 0.04207523648896385
д : 0.028881718853120186
л : 0.04667742172031972
к : 0.03182371489330498
ф : 0.0019153772823934884
т : 0.058970448045757864
у : 0.024273667228862653
с : 0.05518369142773337
е : 0.08809562220429713
ь : 0.017624110874825842
п : 0.02778763657695974
б : 0.016186844613918016
н : 0.06936569626750752
и : 0.07141893378309012
з : 0.01765930923223583
ж : 0.009795409547554448
```

```
h1 з пробілами :  4.396289259217211
h1 без пробілів :  4.4624805277009205
R для h1 без пробілів:  0.115358470787278 79
R для h1 з пробілами :  0.13585820744063848
```

Частота перехресних біграм з пробілами

```
`г : 0.0024597573271216 49
го : 0.006919852389726882
ов : 0.006855285311983153
ва : 0.00507844899945863
ар : 0.0020164795049195145
рд : 0.000347668880158537
д` : 0.0010454899896196006
`л : 0.002477140771129576
ла : 0.003566089370768994
ав : 0.0026571835840688185
вк : 0.00020115128066315355
кр : 0.0018091198513963873
ра : 0.007604015078896002
аф : 0.00017880113836724762
фт : 3.8491911731838025e-05
т` : 0.0036443148688046646
`а : 0.0018612701834201679
вг : 2.731684058388505e-05
гу : 0.00048052805936197793
ус : 0.001136132233375219
```

Частота перехресних біграм без пробілів

```
го : 0.008237894313843681
ов : 0.010666584537045592
ва : 0.006084924961391745
ар : 0.0026618796830383267
рд : 0.0004370469121462377
дл : 0.0007025015802619055
ла : 0.0042736734966246194
ав : 0.004360202918828069
вк : 0.000913692034453376
кр : 0.0022233661705157596
ра : 0.008996126708406113
аф : 0.00032558528354518373
фт : 4.546461166621935e-05
та : 0.006849023757459496
вг : 0.00027425427037364576
гу : 0.0005983729535424999
ус : 0.0018479164741753672
ст : 0.014858128412595751
тд : 0.00038424929859837
де : 0.0049747084765101954
```

```
h2 з пробілами з перетином:  3.991987917987248
h2 без пробілів з перетином:  4.135545001156745
R для h2 з пробілами з перетином:  0.21532833898671933
R для h2 без пробілу з перетином:  0.18017012483499117
```

Частота не перехресних біграм з пробілами

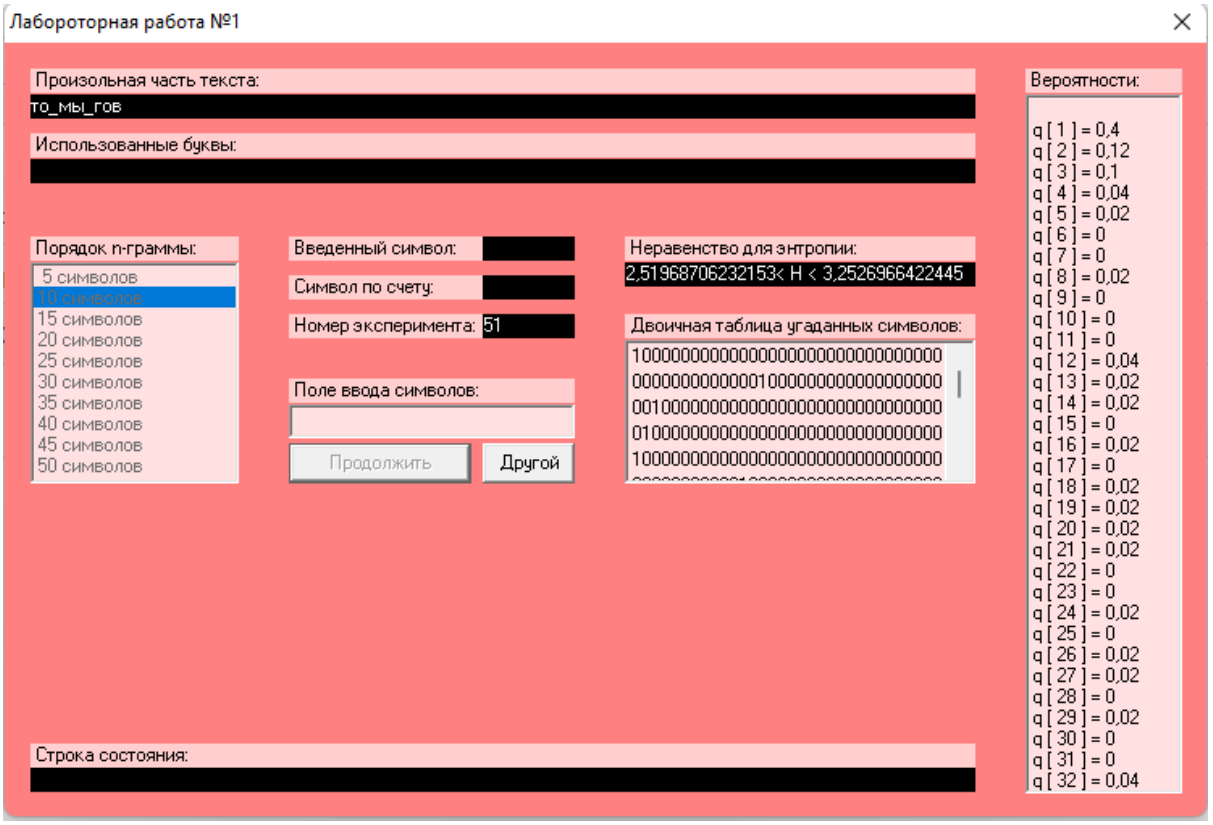
```
`г : 0.002523082730293383
ов : 0.006931027460874834
ар : 0.0019469457288878072
д` : 0.0010057564033157677
ла : 0.003797040841160022
вк : 0.00020115128066315355
ра : 0.007685965600647658
фт : 4.221693544782235e-05
ав : 0.0027068505669486094
гу : 0.0004668696390700354
ст : 0.012262778073020399
`д : 0.0061289056873662096
ер : 0.006392140696629102
ле : 0.004959248240547131
т` : 0.0037175736685523563
`т : 0.006282873334293562
ва : 0.00508093234860262
рь : 0.00044451949677412947
`у : 0.0036107896553608057
`п : 0.01666078940702589
```

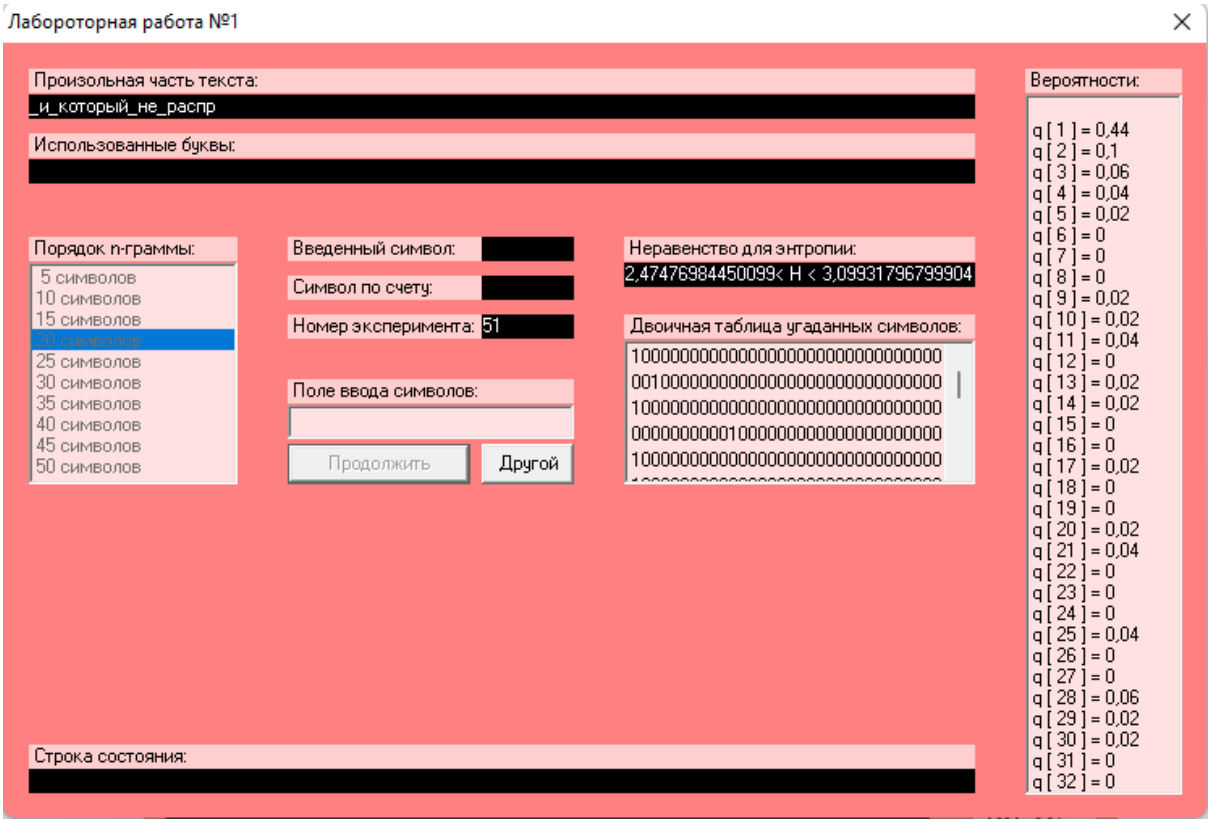
Частота не перехресних біграм без пробілів

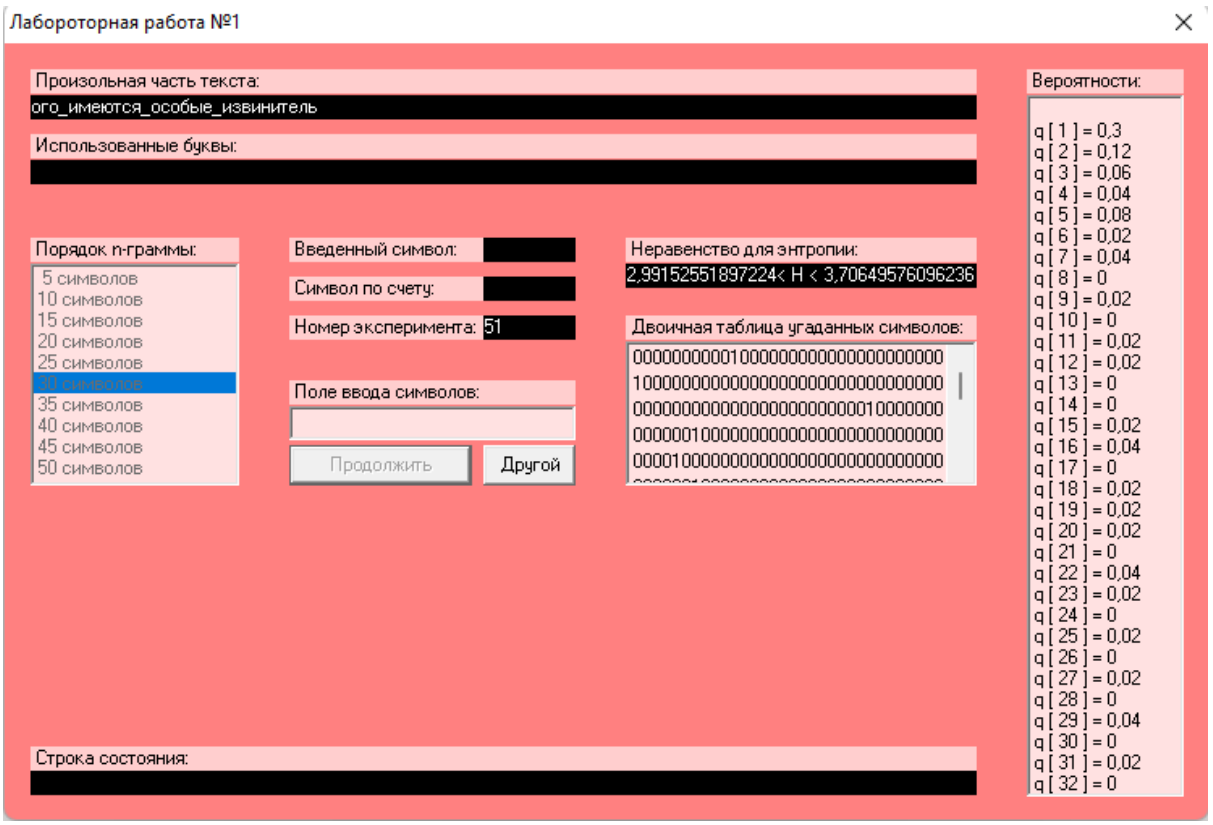
```
го : 0.00823348243748625
ва : 0.006165578939649483
рд : 0.0004877905697734104
ла : 0.00432059837207597
вк : 0.0009210236855613404
ра : 0.009087042604678448
фт : 4.3997946762484416e-05
ав : 0.004455525408814255
гу : 0.0006189044511256142
ст : 0.015061963775023833
де : 0.00503629830607905
рл : 9.092908997580113e-05
ет : 0.00557307325658136
тв : 0.003904084476057784
ар : 0.0026662755738065557
ьу : 0.0002111901444599252
по : 0.011357336657622645
ро : 0.007916697220796362
га : 0.0012788736525628805
сб : 0.00012906064383662097
```

```
h2 з пробілами та без перетину:  3.991402507148115
h2 без пробілів та без перетину:  4.134844212595269
R для h2 з пробілами та без перетину:  0.2154434082967861
R для h2 без пробілів та без перетину:  0.18030904906352985
```

CoolPinkProgramm







H(10)=2,8861
H(20)=2,787
H(30)=3,3489

R(10)=0,42278
R(20)=0,4426
R(30)=0,33022

Висновки

В процесі роботи над цією лабораторною роботою ми дізналися про ентропію та надлишковість, не лише теоретично, а й опрацювали ці поняття на практиці. Також за допомогою програми CoolPinkProgramm ми зрозуміли для чого існує надлишковість, а саме для розуміння тексту при недостатності даних. Ці знання можуть допомогти нам при роботі та в подальшій кар’єрі.