**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: Андреєв Д.Ю.

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.

## Порядок виконання 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. 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

Під час виконання завдань, було використано текст двох книг. Всі маніпуляції з текстом відбуваються за допомогою функції open_sort().

Всі результати виконання роботи були виписані в окремі файли. Далі наведені таблиці з даними підписані та знаходяться нижче.

Частота літер з пробілом

```
 : 0.16392457535500388
о : 0.0947251828908359
е : 0.06919408647436875
а : 0.06469858661696723
и : 0.05752488840840204
н : 0.05657246323943942
т : 0.05387834691887805
с : 0.044281137869512446
л : 0.04162814296989514
в : 0.03667765448720925
р : 0.03565625145085659
к : 0.028222559742128894
м : 0.02756727000192342
д : 0.025552320375664076
у : 0.02384644465521015
п : 0.023360946588580185
ы : 0.017403646541489525
я : 0.017118449589780663
ь : 0.01640479396178361
б : 0.014647184840787144
з : 0.014259847585443018
г : 0.01394148819748894
ч : 0.012463770038402102
й : 0.009394254939544879
ж : 0.008631518905904903
х : 0.007326245415293189
ш : 0.0065157554734601025
ю : 0.004685188992724161
щ : 0.0030509441345599013
ц : 0.002830745557891665
э : 0.0027020753052602255
ф : 0.0011434408017350587
ъ : 0.00016979167357550755
```

Частота літер без пробілу

```
о : 0.11329741324600817
е : 0.08276057929073694
а : 0.07738367222603
и : 0.06880346762115097
н : 0.06766430584113929
т : 0.0644419693854238
с : 0.05296308988906658
л : 0.04978993729850481
в : 0.043868834564516 54
р : 0.042647170817150255
к : 0.03375599725839616
м : 0.032972228568527405
д : 0.030562219175995737
у : 0.028521882060264514
п : 0.02794119513619171
ы : 0.020815880994085243
я : 0.020474767090599852
ь : 0.019621189043738736
б : 0.01751897545249156
з : 0.01705569517426954
г : 0.01667491686340213
ч : 0.01490747087045922
й : 0.011236133323179244
ж : 0.010323851953392735
х : 0.00876266087883634
ш : 0.007793262762419719
ю : 0.005603787474932094
щ : 0.003649125479146041
ц : 0.0033857538141294137
э : 0.003231855913487168
ф : 0.001367628766532125
ъ : 0.00020308176579595 36
```

Частота перехресних біграм з пробілами

```
о` : 0.021283651582179833
е` : 0.018275155366013807
и` : 0.01801914135820074
`п : 0.01655070868126256
`с : 0.015325025037639365
`н : 0.01531043356569147
`в : 0.015126050420168067
а` : 0.01468034727703236
то : 0.014116585860863682
ст : 0.011777970856851028
`о : 0.011055029746705311
я` : 0.010487288838187208
ь` : 0.010241886809972608
но : 0.01005219767464997
на : 0.00975771524079245
`и : 0.00970863483514953
по : 0.009286808646110378
не : 0.00922578976341918
`к : 0.008514787130321742
м` : 0.008310506523051209
`т : 0.007961637693751534
ли : 0.007937760739654978
ал : 0.007924495765156891
ко : 0.007761336578830427
ра : 0.007692358711440377
ни : 0.0074045087648319
й` : 0.007185636685613472
ен : 0.007139209274870169
`д : 0.007107373336074761
он : 0.007052986940632606
ка : 0.006891154251755951
`б : 0.006802278922618715
от : 0.00663116075159345 5
го : 0.0066059573000470904
ло : 0.006552897402054745
ов : 0.006544938417355893
ро : 0.006534326437757423
пр : 0.006477287047415651
`м : 0.006473307555066226
л` : 0.0063539227845834465
во : 0.006343310804984977 6
ть : 0.006006380452733579
та : 0.005896281164399462
ол : 0.005869751215403288
т` : 0.005745060455121275
у` : 0.005729142485723571
ос : 0.005723836495924337
в` : 0.005672103095381799
```

Частота перехресних біграм без пробілами

```
то : 0.0173079608052192
ст : 0.014339476556748661
но : 0.012378468255781485
на : 0.011720039093239916
не : 0.011134592440281268
по : 0.01111872667732846
он : 0.010885499961922169
ен : 0.010468230396263295
ов : 0.01035082375041251
ли : 0.009935140761048917
ал : 0.00974316502931993
ко : 0.009606719467925774
от : 0.009335414921432741
ос : 0.00924180692001117
ра : 0.009213248546696114
ни : 0.0091894499022669
ло : 0.008651600538166678
ка : 0.008297794024319042
во : 0.008159761886629605
го : 0.00800745056228264
ро : 0.007871005000888482
пр : 0.007771050694285787
ом : 0.007734559439494327
ол : 0.007542583707765339
ин : 0.007488640113725789
ес : 0.007406138146371183
ть : 0.007184017465031858 5
та : 0.007157045668012083
ел : 0.007093582616200847 5
ет : 0.007053918208818826
од : 0.007052331632523545
ер : 0.006985695428121747
ва : 0.006790546543802198
ре : 0.006706963115274287 3
ак : 0.006544627218033660 5
те : 0.006460538674383774
ат : 0.006397075622572538
ор : 0.006079760363516361
ас : 0.005930622191759958
ла : 0.005905236971035463
ан : 0.005883024902901531
об : 0.00588143832660625
ем : 0.005868745716244002
ил : 0.005789416901479959
ит : 0.005727540425964003
ис : 0.005640278729723555
за : 0.005583161983093443
де : 0.005572055949026476 6
```

Частота не перехресних біграм без пробілами

```
о` : 0.02143086437112166
е` : 0.018125237111104624
и` : 0.018045647369691695
`п : 0.01647242348109611
`с : 0.015421838894445431
`н : 0.015318372230608623
`в : 0.015015931213239488
а` : 0.014742673101055094
то : 0.013917592781741052
ст : 0.011975603091265557
`о : 0.011161134737473238
я` : 0.010434215099235143
ь` : 0.010306871512974455
но : 0.010089326219779112
на : 0.009678112555812307
`и : 0.009550768969551617
по : 0.009410160426388774
не : 0.009179350176291278
`к : 0.008513449339803095
м` : 0.008433859598390166
ал : 0.008102235675836289
`т : 0.008027951917184221
ли : 0.007903261322303963
ра : 0.007855507477456206
ко : 0.007847548503314913
ни : 0.0073487861237938835
й` : 0.007216136554772334
он : 0.007157770744402852
ен : 0.007147158778881128
`д : 0.006993285278816129
`б : 0.006900430580501044
ка : 0.0068553297270337165
от : 0.006637784433838375
ро : 0.0065926835803710475
го : 0.0065661536665667375
ов : 0.0065343177700015655
`м : 0.006465339994110359
во : 0.006452075037208204
пр : 0.006385750252697429
ло : 0.006377791278556136
л` : 0.006364526321653981
ть : 0.006022290433578381
та : 0.005945353683545882
ол : 0.005762297278296143
в` : 0.00575433830415485
ос : 0.005688013519644075
ел : 0.005648218648937609
ре : 0.005645565657557179
```

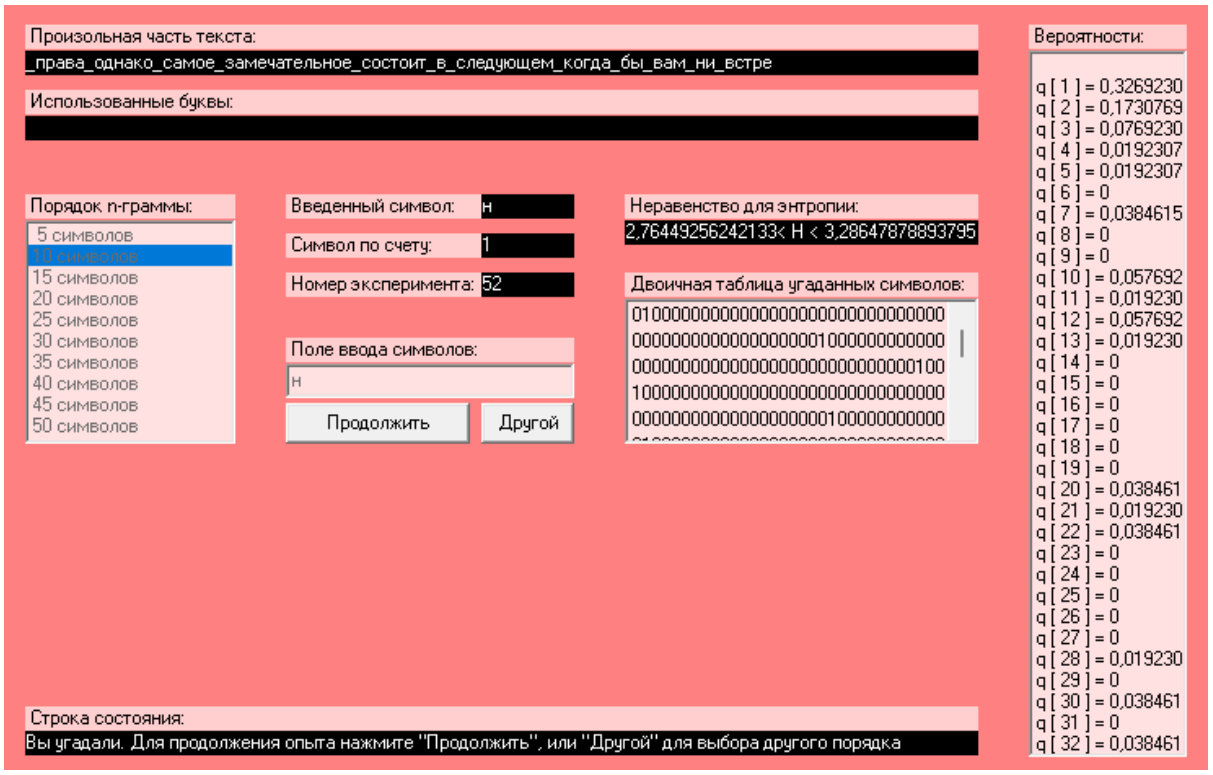
Частота не перехресних біграм з пробілами

```
то : 0.017255603787474934
ст : 0.014431497981874953
но : 0.012286446830655193
на : 0.011575660650369355
по : 0.011052090472926662
не : 0.011001320031477673
он : 0.01090295230117026
ов : 0.010563424973980148
ен : 0.010480923006625542
ал : 0.009858985098875435
ли : 0.009827253572969817
ко : 0.009786002589292514
ос : 0.009478206788008021
от : 0.009392531668062854
ра : 0.009192623054857463
ни : 0.008929251389840834
ло : 0.008494529484933871
ка : 0.008427893280532074
во : 0.008262889345822862
го : 0.008005863985987358
ро : 0.007828167440915898
ом : 0.007790089609829158
ол : 0.00777739699946691
пр : 0.007729799710608483
ин : 0.00750767902926916
ес : 0.007323636179016576
ел : 0.007149112786535679
ет : 0.007126900718401746
од : 0.007117381260630061
ть : 0.007101515497677252
та : 0.007069783971771634
ва : 0.006831797527479502
ре : 0.006784200238621075
ер : 0.006768334475668266
те : 0.006495443352879954
ак : 0.006447846064021527
ат : 0.006327266265580179
ор : 0.00600677785393344
ла : 0.005994085243571192
ан : 0.005911583276216587
ас : 0.005876678597720407
ил : 0.0058735054451298455
ем : 0.005838600766633666
ит : 0.005835427614043104
об : 0.005832254461452542
за : 0.0055561901860736668
ис : 0.005546670728301982
ог : 0.00552445866016805
```

## Ентропія та надлишковість

```
h1 with space: 4.36997
R for h1 with space: 0.14103

h1 without space: 4.45696
R for h1 without space: 0.11645

======================================================================

h2 with space(slide): 3.79578
R for h2 with space(slide): 0.2539

h2 without space(slide_with_no_spaces): 3.85323
R for h2 without space(slide_with_no_spaces): 0.23614

h2 with space(block): 3.56957
R for h2 with space (block): 0.29836

h2 without space(block_with_no_spaces): 3.605
R for h2 without space(block_with_no_spaces): 0.28535
```

# CoolPinkProgram





**Результати**

2.7644 < H(10) <  3.2864

2.6828 < H(20) < 3.2897

2.5846 < H(30) < 3.2472

H(10)  H=3.0254  R=0.39492

H(20)  H=2.9863  R=0.40274

H(30)  H=2.9159  R=0.41682

**Висновок**

Під час роботи ми навчилися рахувати ентропію, та надлишковість російської мови на прикладі вибраного тексту. При виконанні робіт використовується Pyhon3 та програма CoolPinkProgram. Ми отримали знання і закріпили їх на практиці, ці знання будуть нами використовуватися у подальшому житті на роботі чи для особистих дій.