МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

ФІЗИКО- ТЕХНІЧНИЙ ІНСТИТУТ

TO I		• •	••	_
Кафедра	THOO	пмашин	M	оезпеки

Лабораторна робота №1		
з дисципліни Криптографія		

з теми:

«Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Виконала:

Бородай Ю.

Групи ФБ-96

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.

- 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 Н1 та Н2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення Н1 та Н2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення Н1 та Н2 на тому ж тексті, в якому вилучено всі пробіли.
- 2. За допомогою програми CoolPinkProgram оцінити значення (10) H, (20) H, (30) H.
- 3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Як текст я обрала книгу Гаррі Поттер 2гу та 3ю частину

Частота

- o 0.109085
- a 0.083076
- e 0.076115
- и 0.069551
- н 0.063367
- л 0.057628
- p 0.057233
- т 0.054777
- c 0.052384
- в 0.039565
- к 0.032938
- д 0.031340
- м 0.030743
- y 0.030510

- п 0.030458
- г 0.023142
- я 0.020345
- з 0.019640
- ь 0.018663
- ы 0.016280
- б 0.015642
- ч 0.014344
- й 0.009976
- ж 0.009066
- ш 0.008752
- x 0.007671
- ю 0.005064
- ф 0.003571
- ц 0.002941
- э 0.002900
- щ 0.002895
- ъ 0.000283
- ë 0.000046
 - Частота
 - 0.168005
- o 0.090758
- a 0.069119
- e 0.063327
- и 0.057866
- н 0.052721
- л 0.047947
- p 0.047617
- т 0.045574
- c 0.043583

в 0.032917 к 0.027404 д 0.026075 м 0.025578 y 0.025384 п 0.025341 г 0.019254 я 0.016927 з 0.016341 ь 0.015528 ы 0.013545 б 0.013014 ч 0.011934 й 0.008300 ж 0.007543 ш 0.007281 x 0.006382 ю 0.004213 ф 0.002971 ц 0.002447 э 0.002413 щ 0.002409 ъ 0.000235 ë 0.000038 частоти появ біграм з кроком 1 без пробілу a б В ... Ю Я $a \ 0.000307 \ 0.001601 \ 0.00468 \ \dots \ 0.000218 \ 0.000756 \ 0.002365$ $6 \ 0.000884 \ 0.000272 \ 0.00006 \ ... \ 0.00004 \ 0.000039 \ 0.000648$ в 0.006285 0.000223 0.000405 ... 0.000126 0.000002 0.000259

 $\Gamma \ 0.005547 \ 0.000068 \ 0.000279 \ \dots \ 0.000005 \ 0.000005 \ 0.000002$

```
д 0.004803 0.000105 0.001328 ... 0.000042 0.000054 0.000665
```

- e 0.000191 0.002232 0.003512 ... 0.000153 0.000168 0.000473
- ë 0 0.000004 0.000002 ... 0.000001 0
- ж 0.001585 0.000024 0.000041 ... 0.000005 0.000001 0.000008
- $\hbox{ 3 } 0.007769 \ 0.000204 \ 0.000973 \ \dots \ 0.000009 \ 0.000001 \ 0.000286 \\$
- и 0.000334 0.001252 0.004502 ... 0.000226 0.000294 0.001445
- й 0.000108 0.000213 0.000652 ... 0.000049 0.000005 0.000064
- л 0.007326 0.000339 0.000958 ... 0.000493 0.001294 0.001958
- м 0.004194 0.000663 0.000736 ... 0.000101 0 0.000484
- н 0.011913 0.00024 0.000558 ... 0.000047 0.000223 0.001595
- $\hbox{ o } 0.000161 \ 0.004833 \ 0.008594 \ \dots \ 0.000317 \ 0.000173 \ 0.001055 \\$
- п 0.002451 0.000001 0.000011 ... 0.000001 0.000016 0.000358
- p 0.008256 0.000142 0.000809 ... 0.000084 0.000331 0.001299
- $c\ 0.001542\ 0.000235\ 0.001456\ ...\ 0.0002\ 0.00021\ 0.005413$
- $\mathtt{T} \ \ 0.005834 \ \ 0.000251 \ \ 0.002529 \ \dots \ \ 0.000158 \ \ 0.000045 \ \ 0.000783$
- y 0.000427 0.000904 0.001375 ... 0.000108 0.001007 0.000125
- Ф 0.00013 0.000002 0.000014 ... 0.000001 0 0.000001
- x 0.000463 0.00008 0.000622 ... 0.00004 0 0.000015
- ч 0.002436 0.000004 0.000016 ... 0.000001 0 0.000001
- ш 0.001097 0.000002 0.000053 ... 0.000001 0.000001 0.000001
- щ 0.000413 0.000002 0.000005 ... 0 0 0.000004
- ъ 0 0 0 ... 0 0 0.000215
- ы 0.000057 0.000406 0.001366 ... 0.000042 0 0.000087
- ь 0.00013 0.000384 0.001232 ... 0.000113 0.000307 0.000429
- \bullet 0.000001 0.000001 0.000001 ... 0.00008 0 0.000002
- ю 0.000031 0.000284 0.000206 ... 0.000016 0.00004 0.000031
- я 0.000159 0.00046 0.00184 ... 0.000106 0.000092 0.000207

б В ... Ю a 0.000307 0.001601 0.00468 ... 0.000218 0.000756 0.002365 $6 \ 0.000884 \ 0.000272 \ 0.00006 \ \dots \ 0.00004 \ 0.000039 \ 0.000648$ в 0.006285 0.000223 0.000405 ... 0.000126 0.000002 0.000259 $\ \, \Gamma \ \, 0.005547 \ \, 0.000068 \ \, 0.000279 \ \, ... \ \, 0.000005 \ \, 0.000005 \ \, 0.000002$ д 0.004803 0.000105 0.001328 ... 0.000042 0.000054 0.000665 e 0.000191 0.002232 0.003512 ... 0.000153 0.000168 0.000473 $0\ 0.000004\ 0.000002\ \dots\ 0.000001$ 0 0 ж 0.001585 0.000024 0.000041 ... 0.000005 0.000001 0.000008 3 0.007769 0.000204 0.000973 ... 0.000009 0.000001 0.000286 и 0.000334 0.001252 0.004502 ... 0.000226 0.000294 0.001445 й 0.000108 0.000213 0.000652 ... 0.000049 0.000005 0.000064 $\kappa \ 0.007475 \ 0.000201 \ 0.000659 \ ... \ 0.000086$ 0 0.000064 л 0.007326 0.000339 0.000958 ... 0.000493 0.001294 0.001958м 0.004194 0.000663 0.000736 ... 0.0001010 0.000484 н 0.011913 0.00024 0.000558 ... 0.000047 0.000223 0.001595 o 0.000161 0.004833 0.008594 ... 0.000317 0.000173 0.001055 π 0.002451 0.000001 0.000011 ... 0.000001 0.000016 0.000358 p 0.008256 0.000142 0.000809 ... 0.000084 0.000331 0.001299 $c \ 0.001542 \ 0.000235 \ 0.001456 \ \dots \quad 0.0002 \ 0.00021 \ 0.005413$ т 0.005834 0.000251 0.002529 ... 0.000158 0.000045 0.000783 y 0.000427 0.000904 0.001375 ... 0.000108 0.001007 0.000125 ф 0.00013 0.000002 0.000014 ... 0.000001 0 0.000001 x 0.000463 0.00008 0.000622 ... 0.00004 0 0.000015 ц 0.000766 0.000008 0.000088 ... 0.000028 0 0.000002 ч 0.002436 0.000004 0.000016 ... 0.000001 0 0.000001 ш 0.001097 0.000002 0.000053 ... 0.000001 0.000001 0.000001

ш 0.000413 0.000002 0.000005 ...

0 ...

0

0 0.000004

0

0 0.000215

- ы 0.000057 0.000406 0.001366 ... 0.000042 0 0.000087
- ь 0.00013 0.000384 0.001232 ... 0.000113 0.000307 0.000429
- \circ 0.000001 0.000001 0.000001 ... 0.00008 0 0.000002
- ю 0.000031 0.000284 0.000206 ... 0.000016 0.00004 0.000031
- я 0.000159 0.00046 0.00184 ... 0.000106 0.000092 0.000207

частоти появ біграм з кроком 1 з пробілом

- а б в... ь э ю я
- 0 0 0 0 ... 0 0 0
- a 0 0.000307 0.001601 0.00468 ... 0 0.000218 0.000756 0.002365
- $6 \ 0 \ 0.000884 \ 0.000272 \ 0.00006 \ \dots \ 0.00004 \ 0.00004 \ 0.000039 \ 0.000648$
- в 0 0.006285 0.000223 0.000405 ... 0.000182 0.000126 0.000002 0.000259
- $\ \, \Gamma \ \, 0 \ \, 0.005547 \ \, 0.000068 \ \, 0.000279 \ \, \dots \qquad 0 \ \, 0.000005 \ \, 0.000005 \ \, 0.000002$
- д 0 0.004803 0.000105 0.001328 ... 0.000546 0.000042 0.000054 0.000665
- e 0 0.000191 0.002232 0.003512 ... 0 0.000153 0.000168 0.000473
- ж 0 0.001585 0.000024 0.000041 ... 0.00002 0.000005 0.000001 0.000008
- 3 0 0.007769 0.000204 0.000973 ... 0.000191 0.000009 0.000001 0.000286
- и 0 0.000334 0.001252 0.004502 ... 0 0.000226 0.000294 0.001445
- й 0 0.000108 0.000213 0.000652 ... 0 0.000049 0.000005 0.000064
- $\ \kappa \ 0 \ 0.007475 \ 0.000201 \ 0.000659 \ \dots \ 0.000004 \ 0.000086 \qquad \ \ 0 \ 0.000064$
- л 0 0.007326 0.000339 0.000958 ... 0.005135 0.000493 0.001294 0.001958
- м 0 0.004194 0.000663 0.000736 ... 0.000109 0.000101 0 0.000484
- н 0 0.011913 0.00024 0.000558 ... 0.001229 0.000047 0.000223 0.001595
- o 0 0.000161 0.004833 0.008594 ... 0 0.000317 0.000173 0.001055
- п 0 0.002451 0.000001 0.000011 ... 0.000025 0.000001 0.000016 0.000358
- p 0 0.008256 0.000142 0.000809 ... 0.000637 0.000084 0.000331 0.001299
- $c \ 0 \ 0.001542 \ 0.000235 \ 0.001456 \ \dots \ 0.004123 \ \ 0.0002 \ \ 0.00021 \ \ 0.005413$
- т 0 0.005834 0.000251 0.002529 ... 0.005335 0.000158 0.000045 0.000783
- y 0 0.000427 0.000904 0.001375 ... 0 0.000108 0.001007 0.000125

```
ф 0 0.00013 0.000002 0.000014 ... 0.000001 0.000001
                                                       0 0.000001
x 0 0.000463 0.00008 0.000622 ...
                                   0 0.00004
                                                    0 0.000015
ц 0 0.000766 0.000008 0.000088 ... 0 0.000028
                                                    0 0.000002
ч 0 0.002436 0.000004 0.000016 ... 0.000263 0.000001
                                                        0 0.000001
_{
m III} 0 0.001097 0.000002 0.000053 ... 0.000787 0.000001 0.000001 0.000001
ш 0 0.000413 0.000002 0.000005 ... 0.000036
                                                     0 0.000004
              0
                   0 ...
                                  0
                                        0 0.000215
ы 0 0.000057 0.000406 0.001366 ...
                                     0 0.000042
                                                     0 0.000087
ь 0 0.00013 0.000384 0.001232 ... 0 0.000113 0.000307 0.000429
_{\odot} 0 0.000001 0.000001 0.000001 ...
                                    0 0.00008
                                                    0 0.000002
ю 0 0.000031 0.000284 0.000206 ... 0 0.000016 0.00004 0.000031
```

я 0 0.000159 0.00046 0.00184 ... 0 0.000106 0.000092 0.000207

частоти появ біграм з кроком 2 з пробілом

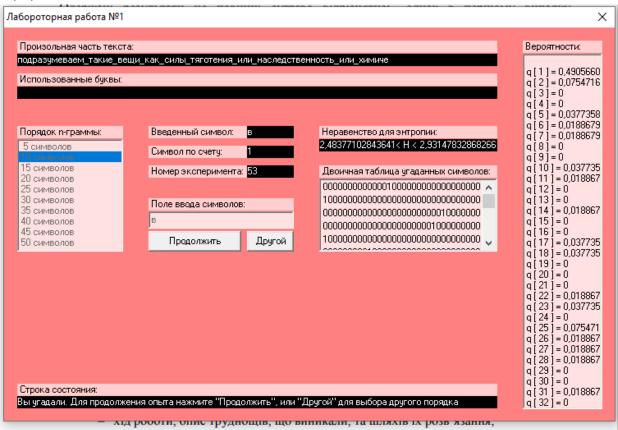
Ю Я 0 0 0 ... 0 0 0 0 a 0 0.000307 0.001601 0.00468 ... 0 0.000218 0.000756 0.002365 $6 \ 0 \ 0.000884 \ 0.000272 \ 0.00006 \ \dots \ 0.00004 \ 0.00004 \ 0.000039 \ 0.000648$ в 0 0.006285 0.000223 0.000405 ... 0.000182 0.000126 0.000002 0.000259 Γ 0 0.005547 0.000068 0.000279 ... 0 0.000005 0.000005 0.000002 д 0 0.004803 0.000105 0.001328 ... 0.000546 0.000042 0.000054 0.000665 e 0 0.000191 0.002232 0.003512 ... 0 0.000153 0.000168 0.000473 0 0.000004 0.000002 ... 0 0.000001 ж 0 0.001585 0.000024 0.000041 ... 0.00002 0.000005 0.000001 0.000008 $3\ 0\ 0.007769\ 0.000204\ 0.000973\ \dots\ 0.000191\ 0.000009\ 0.000001\ 0.000286$ и 0 0.000334 0.001252 0.004502 ... 0 0.000226 0.000294 0.001445 й 0 0.000108 0.000213 0.000652 ... 0 0.000049 0.000005 0.000064 к 0 0.007475 0.000201 0.000659 ... 0.000004 0.000086 0 0.000064 л 0 0.007326 0.000339 0.000958 ... 0.005135 0.000493 0.001294 0.001958 м 0 0.004194 0.000663 0.000736 ... 0.000109 0.000101 0 0.000484

```
н 0 0.011913 0.00024 0.000558 ... 0.001229 0.000047 0.000223 0.001595
o 0 0.000161 0.004833 0.008594 ...
                                     0 0.000317 0.000173 0.001055
\pi 0 0.002451 0.000001 0.000011 ... 0.000025 0.000001 0.000016 0.000358
p 0 0.008256 0.000142 0.000809 ... 0.000637 0.000084 0.000331 0.001299
c\ 0\ 0.001542\ 0.000235\ 0.001456\ ...\ 0.004123\ 0.0002\ 0.00021\ 0.005413
т 0 0.005834 0.000251 0.002529 ... 0.005335 0.000158 0.000045 0.000783
y 0 0.000427 0.000904 0.001375 ...
                                     0 0.000108 0.001007 0.000125
ф 0 0.00013 0.000002 0.000014 ... 0.000001 0.000001
                                                      0 0.000001
x 0 0.000463 0.00008 0.000622 ...
                                    0 0.00004
                                                  0 0.000015
ц 0 0.000766 0.000008 0.000088 ...
                                     0 0.000028
                                                   0 0.000002
ч 0 0.002436 0.000004 0.000016 ... 0.000263 0.000001
                                                      0 0.000001
щ 0 0.000413 0.000002 0.000005 ... 0.000036
                                              0
                                                    0 0.000004
ъ 0
                   0 ...
                                       0 0.000215
ы 0 0.000057 0.000406 0.001366 ...
                                     0 0.000042
                                                    0 0.000087
ь 0 0.00013 0.000384 0.001232 ...
                                    0\;\; 0.000113\;\; 0.000307\;\; 0.000429
\bullet 0 0.000001 0.000001 0.000001 ...
                                    0.00008
                                                   0 0.000002
ю 0 0.000031 0.000284 0.000206 ...
                                     0 0.000016 0.00004 0.000031
я 0 0.000159 0.00046 0.00184 ...
                                   0 0.000106 0.000092 0.000207
```

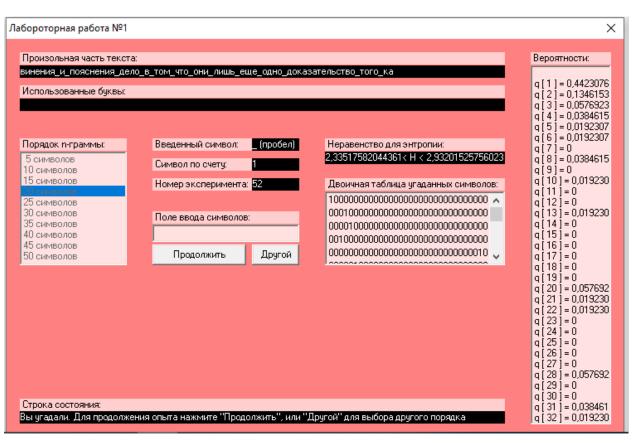
ентропія Н1 без пробілу 4.463763716161964
ентропія Н1 з пробілом 4.366965455558461
ентропія Н2 біграм без пробілу 4.1688180862464055
ентропія Н2 біграм з кроком 2 без пробілу 4.16818610720163
ентропія Н2 біграм з пробілом 4.005939898826621
ентропія Н2 біграм з кроком 2 з пробілом 4.006355973640299
ентропія мови 0.13526556715642313
Н0 5.044394119358453

надлишковість 0.9731849724752224

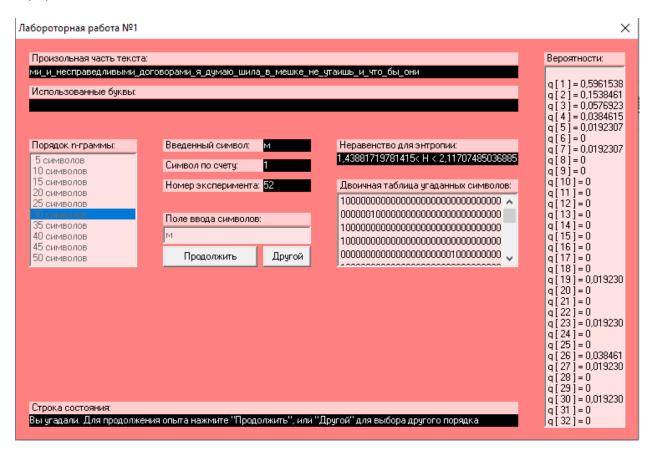
H(10)



H(20)



H(30)



Висновки: в цьому лабораторному практикумі я отримала навички порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела. Засвоїла поняття ентропії на символ джерела та його надлишковості