

НТУУ "КПІ ім Ігоря Сікорського"

Фізико-технічний інститут

КРИПТОГРАФІЯ

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела
відкритого тексту

Виконали:

студенти групи ФБ-14

Гавриленко Давид

Земляний Олександр

Перевірила:

Селюх П.В.

Київ 2023

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 H_1 та H_2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H_1 та H_2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 на тому ж тексті, в якому вилучено всі пробіли
Обраний текст міститься у файлі `avidreaders.ru__kobzar.txt`
Програма `main.py` на мові Python3

Букви без пробілів:

```
1 к: 0.034564282366467365
2 о: 0.10184655302333892
3 б: 0.02213481313069846
4 з: 0.021982034238008798
5 а: 0.1013364268901209
6 р: 0.04381387905608539
7 т: 0.05901149466957033
8 в: 0.04653282545140984
9 с: 0.04688499374642329
10 ш: 0.008079155003249789
11 е: 0.05797570556658959
12 ч: 0.015249405068633976
13 н: 0.05760800043503142
14 г: 0.018763319600496143
15 я: 0.0241442439904811
16 у: 0.0360092081651255
17 ь: 0.02097990828087493
18 д: 0.035923755564129585
19 и: 0.07121309030268347
20 й: 0.017243299091871903
21 п: 0.028810473899409342
22 ю: 0.012486437636432845
23 м: 0.03450990343856088
24 л: 0.047685140828475914
25 ц: 0.005119387641482318
26 щ: 0.005184124460418614
27 х: 0.01203069043112132
28 ф: 0.0003469893494985486
29 ы: 0.0019369256225739877
30 э: 5.6968400663940814e-05
31 ж: 0.010523617286284339
32 ъ: 1.2947363787259277e-05
```

Букви з пробілами:

```
1 к: 0.02946780022473944
2 о: 0.08682934152226753
3 б: 0.018871048570652738
4 з: 0.018740796831571257
5 а: 0.08639443317313106
6 р: 0.03735354958065563
7 _: 0.14744938395342727
8 т: 0.050310286134371225
9 в: 0.039671589004987094
10 с: 0.03997183030185288
11 ш: 0.006887888575156357
12 е: 0.04942722349653067
13 ч: 0.013000889685607624
14 н: 0.04911373626009727
15 г: 0.01599667968448172
16 я: 0.02058419008806342
17 у: 0.03069967260452702
18 ь: 0.017886433729460516
19 д: 0.030626819936905175
20 и: 0.060712764008133004
21 й: 0.0147007852634507
22 п: 0.02456238727153514
23 ю: 0.010645320099167934
24 м: 0.029421439436252812
25 л: 0.04065399618958472
26 ц: 0.004364537087526961
27 щ: 0.0044197285023919956
28 х: 0.010256772538518089
29 ф: 0.0002958259836765871
30 ы: 0.0016513271327618447
31 э: 4.856844508123073e-05
32 ж: 0.008971916400460076
33 ь: 1.1038282973006984e-05
```

Біграми з пробілами без повторень:

```
1 ко: 0.007024025460979491
2 об: 0.004155214048050922
3 бз: 0.00016914636724792183
4 за: 0.00668350711638828
5 ар: 0.0045892080166475635
6 р_: 0.0016647563513348096
7 _з: 0.00910719651024337
8 зб: 0.0002715244316348219
9 бр: 0.0021165552006943904
10 рк: 0.0005385976430789091
```

Біграми без пробілів без повторень:

```
1 ko: 0.008377820411227426
2 об: 0.005859782291362556
3 бз: 0.00037014639029069523
4 за: 0.007835634149393169
5 ар: 0.006013575509863619
6 рз: 0.00040663969637569336
7 зб: 0.0004952662968678317
8 бр: 0.0025284647787462986
9 рк: 0.0007011928097760354
10 ка: 0.007095341368811778
```

Біграми із пробілами із повтореннями:

```
1 /home/david/Pictures/Screen
2 об: 0.004121703961427739
3 бз: 0.0001677822715953445
4 за: 0.006629607389484467
5 ар: 0.004552197947757899
6 р_: 0.0016513307783331272
7 __: 0.004938538704720863
8 _з: 0.009033750728528285
9 зб: 0.0002693346991398951
10 бр: 0.002099486056410166
```

Біграми без пробілів із повтореннями:

```
1 ko: 0.008322586993562553
2 об: 0.005821149832460679
3 бз: 0.00036770608372305
4 за: 0.007783975265292171
5 ар: 0.005973929120768143
6 рз: 0.0004039587962027873
7 зб: 0.0004920010979392922
8 бр: 0.0025117950789532288
9 рк: 0.0006965699755035243
10 ка: 0.007048563097846071
```

Ентропія та надлишковість:

```
1 H1: 4.421004551966904
2 H2: 8.105516806574938
3 H2 (no repeats): 8.083859383040554
4 H1 (no underscores): 4.477838702781195
5 H2 (no underscores): 8.308692657245142
6 H2 (no repeats, no underscores): 8.280231377200046
7 R1: 0.12358066254165567
8 R2: -0.6068365426620945
9 R2 (no repeats): -0.6025431779840114
10 R1 (no underscores): 0.104432259443761
11 R2 (no underscores): -0.6617385314490285
12 R2 (no repeats, no underscores): -0.6560462754400092
```