

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:  
Студент гр. ФБ-11  
Ахунов Михайло

## 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.

## Літери з пробілом

Символ(и)	Кількість	Частота
	101331	0.15029456435326363
о	61112	0.09064157480688681
а	48228	0.07153197194964225
е	46568	0.06906985298480012
и	41170	0.06106351673647614
н	34964	0.05185875149803624
т	33334	0.04944112865906475
с	30936	0.045884404997804856
л	29330	0.04350237905952988
р	27246	0.04041138151571603
в	27170	0.04029865799684374
к	20669	0.030656347520675867
п	17755	0.02633429049444095
м	17585	0.026082145781173984
д	17369	0.025761773674905373
у	16536	0.024526264579897244
ы	11203	0.016616336604292986
я	10530	0.015618140180594943
ь	10011	0.014848357203032856
г	9736	0.014440476049218647
б	9725	0.014424160803066079
з	9557	0.014174982498190491
ч	8458	0.012544941087129348
й	7131	0.010576729119451332
ж	5812	0.008620382785338823
х	5244	0.007777922802188023

ш	4936	0.00732109590991611
ю	3095	0.004590516985654449
ц	2950	0.004375452377279685
щ	1904	0.0028240207885900067
э	1374	0.0020379225648753515
ф	1247	0.0018495556320229718

Ентропія для літер з пробілом: 4.405542220937608

Надлишковість для літер з пробілом: 0.11889155581247834

### Літери без пробілу

Символ(и)	Кількість	Частота
о	61112	0.10667411435104776
а	48228	0.08418443492149384
е	46568	0.08128682021697199
и	41170	0.07186433577419551
н	34964	0.06103144610174817
т	33334	0.05818619792803093
с	30936	0.054000366565715635
л	29330	0.051197011616642085
р	27246	0.04755928327674839
в	27170	0.04742662139871004
к	20669	0.03607879417335067
п	17755	0.03099225848119605
м	17585	0.030695514806636584
д	17369	0.03031847578484338
у	16536	0.028864431779501994
ы	11203	0.019555408153468845
я	10530	0.018380652312418724
ь	10011	0.01747471132949894
г	9736	0.016994684797123332
б	9725	0.016975483735828308
з	9557	0.016682231163322483
ч	8458	0.01476387058484687
й	7131	0.012447524372256212
ж	5812	0.010145142567880116
х	5244	0.009153669584646133
ш	4936	0.008616039868385453
ю	3095	0.005402480428009112
ц	2950	0.005149375529120155
щ	1904	0.0033235291550660255
э	1374	0.002398387110851218
ф	1247	0.002176702130445028

Ентропія для літер без пробілу: 4.466212760632701

Надлишковість для літер без пробілу: 0.0984990337849706

### Біграми з пробілом, які перетинаються

Символ(и)	Кількість	Частота
и	10757	0.015954827533016126
а	10708	0.015882150527427412
п	10703	0.01587473450644897
о	10671	0.015827271972186956
е	10012	0.014849840407228544
с	9969	0.01478606262681396
в	9627	0.014278806791888653
н	8331	0.012356574154276968
ст	7925	0.011754393250827628
по	6471	0.009597814350297234

Ентропія біграм з пробілом, які перетинаються: 4.040149461523215

Надлишковість біграм з пробілом, які перетинаються: 0.19197010769535705

### Біграми без пробілу, які перетинаються

Символ(и)	Кількість	Частота
ст	8038	0.014030739153582307
ов	6671	0.011644570899918832
то	6613	0.01154332894036325
по	6507	0.011358300531520288
но	6477	0.011305934000715676
на	6158	0.01074910322315997
ал	6121	0.010684517835167617
ко	5814	0.010148633669933757
ли	5758	0.010050882812431815
ос	5668	0.00989378322001798

Ентропія біграм без пробілу, які перетинаються: 4.166684715833162

Надлишковість біграм без пробілу, які перетинаються: 0.15895849603348

**Біграми з пробілом, які не перетинаються**

Символ(и)	Кількість	Частота
и	5433	0.01611649679034612
п	5369	0.015926646653298054
о	5321	0.015784259050512003
а	5305	0.015736796516249985
е	5062	0.015015959277145604
с	4996	0.014820176323314784
в	4764	0.014131969576515538
н	4150	0.012310594824210638
ст	3946	0.011705447512369923
то	3301	0.009792114099932366

Ентропія біграм з пробілом, які не перетинаються: 4.039900586024084

Надлишковість біграм з пробілом, які не перетинаються: 0.19201988279518323

**Біграми без пробілу, які не перетинаються**

Символ(и)	Кількість	Частота
ст	4070	0.014208760556201409
то	3360	0.01173008242477561
ов	3319	0.011586947490425669
по	3282	0.011457776939914748
но	3185	0.011119140631818547
ал	3058	0.010675771444929706
на	3055	0.010665298157050442
ко	2944	0.010277786505517677
ен	2831	0.009883292662065403
ли	2822	0.00985187279842761

Ентропія біграм без пробілу, які не перетинаються: 4.166673183722701

Надлишковість біграм без пробілу, які не перетинаються: 0.15896082377944276

2. За допомогою програми CoolPinkProgram оцінити значення  $H(10)$ ,  $H(20)$ ,  $H(30)$ .

[illegible]

Произвольная часть текста:

туго\_с\_деньгами\_а\_т

Использованные буквы:

Порядок n-граммы:

5 символов

10 символов

15 символов

20 символов

25 символов

30 символов

35 символов

40 символов

45 символов

50 символов

Введенный символ:

Символ по счету:

Номер эксперимента:

51

Поле ввода символов:

Продолжить

Другой

Неравенство для энтропии:

1,49896716447422 < H < 2,0992247882468

Двоичная таблица угаданных символов:

00000000000000000000000000000000000010

10000000000000000000000000000000000000

10000000000000000000000000000000000000

00100000000000000000000000000000000000

00000000000000000000000001000000000000

.....

Вероятности:

q[1] = 0,62

q[2] = 0,14

q[3] = 0,04

q[4] = 0

q[5] = 0

q[6] = 0

q[7] = 0

q[8] = 0

q[9] = 0,02

q[10] = 0

q[11] = 0,02

q[12] = 0

q[13] = 0

q[14] = 0,02

q[15] = 0

q[16] = 0,04

q[17] = 0

q[18] = 0,02

q[19] = 0

q[20] = 0

q[21] = 0,02

q[22] = 0

q[23] = 0

q[24] = 0

q[25] = 0

q[26] = 0,02

q[27] = 0

q[28] = 0,02

q[29] = 0

q[30] = 0

q[31] = 0,02

q[32] = 0

Строка состояния:

