

Міністерство освіти і науки України Національний технічний  
університет України "Київський політехнічний інститут імені  
Ігоря Сікорського"

Фізико-технічний інститут

## **КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1**

**Експериментальна оцінка ентропії на символ джерела  
відкритого тексту**

Виконав: Дмитрів А.Ю. ФБ-13

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.

## Хід роботи

### 1. Підрахунок частоти, ентропії та надлишковості для букв

Текст з пробілами

буква	частота
	0.15728132347099022
о	0.09545504548463032
е	0.07221258395314084
а	0.0675637992704478
и	0.056682646254895934
н	0.053630967630643823
т	0.049039196345002686
с	0.044610059429157016
л	0.04126417705729737
р	0.0396371024183548
в	0.03744574133679605
к	0.029054538796337546
м	0.026409263361327154
д	0.025645795499523245
у	0.023935759058460293
п	0.02179994949197774
ы	0.017504209306413752
я	0.01724472525572884
ь	0.016454943518362516
г	0.015680877012234855
б	0.01497442254748284
з	0.014276739374584962
ч	0.01143739910730177
й	0.010708285471926567
ж	0.008498650865671685
ш	0.00781193180759148
х	0.0077717300532600155
ю	0.005216360359754536
ц	0.0038948190537311067
щ	0.0033133554979005546
э	0.0025239392310281526
ф	0.0010196626780435195

Ентропія літер в тексті з пробілами: 4.3919468423307135

Надлишковість: 0.1216106315338573

## Текст без пробілів

буква	частота
о	0.11327035717042681
е	0.08569002439885404
а	0.08017361090029429
и	0.06726164713515126
н	0.0636404165759552
т	0.05819165720520798
с	0.05293588557084905
л	0.0489655423649281
р	0.04703479764043219
в	0.04443445052271487
к	0.03447715068569193
м	0.03133817262731336
д	0.03043221446705461
у	0.02840302431298534
п	0.02586859660185459
ы	0.02077111827936471
я	0.020463205261756424
ь	0.01952602211943035
г	0.018607487230283375
б	0.01776918319783715
з	0.016941287492746697
ч	0.013572025191621644
й	0.012706832980173003
ж	0.010084801847131161
ш	0.009269916551235425
х	0.009222211717521465
ю	0.006189919014539566
ц	0.004621731026269751
щ	0.003931745658643292
э	0.002994996196623712
ф	0.001209968055108624

Ентропія літер в тексті без пробілів: 4.466711351138335

Надлишковість: 0.0983983937468300

## 2. Підрахунок частоти, ентропії та надлишковості для біграм

Текст з пробілами (біграми, що перетинаються):

біграма	частота
ох	0.00053577993990201
бз	3.2892356474202525e-06
уш	0.0008365622663272176
кг	7.309412549822784e-07
уа	0.00047730463950342776
ед	0.0030622783877482553
лы	0.0007002417222730226
э	0.00227139994985743
ви	0.002867482543295478
юр	4.38564752989367e-05
во	0.006981585396963232
да	0.003937946011217024
гл	0.0014907546895363566
ог	0.004971131475134475
лт	8.77129505978734e-05
тч	0.00012535642522946074
лщ	9.867706942260757e-06
жж	1.5715236982118985e-05
кс	0.000212703905199843
ьч	7.345959612571897e-05
йщ	1.827353137455696e-06
хс	8.880936248034681e-05
сэ	4.824212282883037e-05
вы	0.002243258711540612
не	0.007453407977054292
гш	5.116588784875948e-06
ке	0.000616914419205043
рл	0.00010123536381504554
ет	0.004216800099992764
жь	3.6547062749113916e-05
шр	9.136765687278479e-06
як	0.0003497553905090202

Ентропія біграм(перетинаються) в тексті з пробілами: 4.006044507949312  
Надлишковість: 0.19879109841013753

Текст без пробілів (біграми, що перетинаються):

біграма	частота
жа	0.0016909202483254512
ед	0.004492929923737285
ус	0.0022755215550047812
цч	1.0408331869661664e-05
бр	0.001764212251907652
ув	0.001321424466952462
чю	8.67360989138472e-07
эа	8.67360989138472e-07
йж	0.00015569129755035573
эм	1.3444095331646316e-05
шо	0.0002576062137741262
нз	0.000185615251675633
хй	7.372568407677012e-06
мэ	0.00018865101513761766
фс	2.9490273630708048e-05
че	0.003772586622257784
вц	9.844547226721657e-05
кх	5.464374231572374e-05
ут	0.002131973311302364
цг	1.0408331869661664e-05
нб	0.00023201906459454125
яп	0.0011271356053854443
нч	0.0002840607239428496
эу	1.1709373353369372e-05
ик	0.004103918520108681
ыз	0.00028709648740483424
фф	8.239929396815484e-06
ум	0.0013834407776758628
кт	0.0007138380940609624
ни	0.009622502813502208
тс	0.0018791375829684996

Ентропія біграм(перетинаються) в тексті без пробілів :  
4.163605224405998

Надлишковість: 0.15958008856519046

Текст з пробілами (біграми, що не перетинаються):

біграма	частота
жа	0.0014026762683109922
ед	0.0030611819758657815
ус	0.0013376224966175694
бр	0.0014772322763191845
ув	0.0006139906541851138
эм	1.2426001334698732e-05
шо	0.00020685637515998476
нз	3.435423898416708e-05
мэ	5.993718290854682e-05
фс	1.5349766354627843e-05
че	0.003143778337678779
вц	3.508518023914936e-05
кх	3.654706274911392e-06
ут	0.0014896582776538832
цг	2.192823764946835e-06
нб	5.847530039858227e-06
яп	2.558294392437974e-05
нч	0.00017323307743079996
ик	0.002544406508593311
ыз	8.77129505978734e-05
фф	9.502236314769619e-06
ум	0.0008924792723333619
кт	0.00035669933243135185
ни	0.00789562743631857
тс	0.0010971428237283997
гм	4.38564752989367e-06
зб	0.00023901779037920502
сб	6.651565420338733e-05
дл	0.0006943941922331644
бы	0.00328850470616527
йю	7.309412549822784e-07
ши	0.0019070257342487641
бс	0.00011256495326727087
вч	4.8973064083812645e-05
ч	9.502236314769619e-05

Ентропія біграм(не перетинаються) в тексті з пробілами: 4.005971507282249  
Надлишковість: 0.19880569854355024

Текст без пробілів (біграми, що не перетинаються):

біграма	частота
жа	0.0016909202483254512
ед	0.004492929923737285
ус	0.0022755215550047812
цч	1.0408331869661664e-05
бр	0.001764212251907652
ув	0.001321424466952462
чю	8.67360989138472e-07
эа	8.67360989138472e-07
йж	0.00015569129755035573
эм	1.3444095331646316e-05
шо	0.0002576062137741262
нз	0.000185615251675633
хй	7.372568407677012e-06
мэ	0.00018865101513761766
фс	2.9490273630708048e-05
че	0.003772586622257784
вц	9.844547226721657e-05
кх	5.464374231572374e-05
ут	0.002131973311302364
цг	1.0408331869661664e-05
нб	0.00023201906459454125
яп	0.0011271356053854443
нч	0.0002840607239428496
эу	1.1709373353369372e-05
ик	0.004103918520108681
ыз	0.00028709648740483424
фф	8.239929396815484e-06
ум	0.0013834407776758628
кт	0.0007138380940609624
ни	0.009622502813502208
тс	0.0018791375829684996
зб	0.0003317655783454655

Ентропія біграм(не перетинаються) в тексті без пробілів:

4.16347638913816

Надлишковість: 0.15960609384632318

### Результати експерименту для $H^{(10)}$

### Результати експерименту для $H^{(10)}$

[illegible]



### Результати експерименту для $H^{(30)}$

**Лабораторная работа №1**

---

Произвольная часть текста:  
в\_содеянном\_ими\_эле\_не\_более\_

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:  
 $2,22537329819122 < H < 2,88381449563304$

Двоичная таблица угаданных символов:

10000000000000000000000000000000
00000000010000000000000000000000
0000000000000000000000000100000000
00000010000000000000000000000000
00000000000000000100000000000000

Вероятности:

$q[1] = 0,4$
$q[2] = 0,22$
$q[3] = 0,06$
$q[4] = 0,02$
$q[5] = 0$
$q[6] = 0$
$q[7] = 0,06$
$q[8] = 0,04$
$q[9] = 0$
$q[10] = 0,02$
$q[11] = 0,02$
$q[12] = 0$
$q[13] = 0$
$q[14] = 0$
$q[15] = 0,02$
$q[16] = 0$
$q[17] = 0,02$
$q[18] = 0,02$
$q[19] = 0$
$q[20] = 0$
$q[21] = 0$
$q[22] = 0$
$q[23] = 0$
$q[24] = 0,04$
$q[25] = 0$
$q[26] = 0,02$
$q[27] = 0$
$q[28] = 0,02$
$q[29] = 0$
$q[30] = 0$
$q[31] = 0$
$q[32] = 0,02$

Строка состояния:

Результати експериментів у наведених таблицях:

	Найвище значення H	Найнижче значення H
H10	3,6867	3,04821
H20	2,7398	1,9515
H30	2,8838	2,2254
	↓	↓
R10	0,26266	0,390358
R20	0,45204	0,6097
R30	0,42324	0,55492

## Результати з першого та другого завдань

	H	R
літери з пробілами	4.3919468423307135	0.12161063153385732
літери без пробілів	4.466711351138335	0.09839839374683002
біграми(перет.) з пробілами	4.006044507949312	0.19879109841013753
біграми(не перет.) з пробілами	4.005971507282249	0.19880569854355024
біграми(перет.) без пробілів	4.163605224405998	0.15958008856519046
біграми(не перет.) без пробілів	4.16347638913816	0.15960609384632318

**Висновок:** в результаті виконання даного практикуму було на практиці ознайомлено з такими поняттями як частота зустрічі букв та біграм в тексті, а також здійснено обрахунки ентропії та надлишковість деякого джерела тексту(мови). З використанням програми CoolPinkProgram було проведено експерименти, щоб визначити, як кількість букв впливає на ентропію та надлишковість.