

Міністерство освіти і науки України  
Національний технічний університет України  
“Київський політехнічний інститут ім. Ігоря Сікорського”  
Фізико-технічний інститут

## **КРИПТОГРАФІЯ**

### **КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1**

Експериментальна оцінка ентропії на символ  
джерела відкритого тексту

Виконали Студенти:  
Дудченко И.В і Терпило С.Е

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***

Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.

1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку  $1 H$  та  $2 H$  за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення  $1 H$  та  $2 H$  на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення  $1 H$  та  $2 H$  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.

## ***Хід роботи***

Перед виконанням роботи були розглянуті теоретичні відомості в методичних вказівках. В якості експериментального тексту була взята книга «Good Omens» в перекладі на російську мову. Оригінал тексту можна знайти у файлі badtxt.txt. Відредагований текст з пробілами міститься у файлі spaces.txt, а без пробілів у nospaces.txt. В ході виконання роботи було прийнято рішення пробіли замінити на «\_» для кращого сприйняття. Усі таблиці також наведені у файлах з відповідними назвами. Єдина відмінність - таблиці біграм були виконанні у двох варіантах(таблиці у вигляді матриці і звичайні).

## Аналіз результатів для тексту з пробілами

### Текст без пробілів

а	0,077776
б	0,016856
в	0,047941
г	0,020603
д	0,029395
е	0,087385
э	0,003103
ж	0,00878
з	0,015413
и	0,073477
ы	0,015419
й	0,013549
к	0,037569
л	0,042051
м	0,033064
н	0,067137
о	0,109828
п	0,028976
р	0,050127
с	0,056236
т	0,053851
у	0,026052
ф	0,002366
х	0,008574
ц	0,004231
ч	0,015297
ш	0,008547
щ	0,003281
ь	0,014809
ю	0,006976
я	0,021133

### Текст з пробілами

а	0,065903
б	0,014283
в	0,040623
г	0,017458
д	0,024908
е	0,074046
э	0,002629
ж	0,00744
з	0,01306
и	0,062261
ы	0,013065
й	0,01148
к	0,031834
л	0,035632
м	0,028017
н	0,056889
о	0,093063
п	0,024553
р	0,042475
с	0,047651
т	0,045631
у	0,022075
ф	0,002005
х	0,007265
ц	0,003585
ч	0,012962
ш	0,007243
щ	0,00278
ь	0,012548
ю	0,005911
я	0,017907

Біграми

Текст з пробілами

аб	0,000787
вг	0,000202
де	0,004144
эж	0
зи	0,000551
ый	0,001765
кл	0,000498
мн	0,001576
оп	0,001417
рс	0,000719
ту	0,001207
фх	0
цч	0
шщ	0
ью	0,000362

Текст без пробілів

аб	0,001658
вг	0,000987
де	0,004973
эж	0
зи	0,000686
ый	0,002083
кл	0,000688
мн	0,002537
оп	0,003907
рс	0,00119
ту	0,001517
фх	1E-05
цч	8,03E-06
шщ	0
ью	0,000435
я	0

## Перехрестні біграми

`crossed_bigrams(text_with_space)`

а б	0,000787
б в	6,97E-05
в г	0,000202
г д	0,000655
д е	0,004144
е з	1,7E-06
э ж	0
ж з	0
з и	0,000551
и ы	0
ы й	0,001765
й к	8,67E-05
к л	0,000498
л м	3,91E-05
м н	0,001576
н о	0,010083
о п	0,001417
п р	0,005989
р с	0,000719
с т	0,01211
т у	0,001207
у ф	1,36E-05
ф х	0
х ц	1,7E-06
ц ч	0
ч ш	0,000112
ш щ	0
щ ь	2,72E-05
ью	0,000362
ю я	1,7E-06

`crossed_bigrams(text_without_space)`

аб	0,001658
бв	9,43E-05
вг	0,000987
гд	0,000825
де	0,004973
ез	0,000245
эж	0
жз	6,02E-06
зи	0,000686
иы	0
ый	0,002083
йк	0,000777
кл	0,000688
лм	0,000544
мн	0,002537
но	0,012102
оп	0,003907
пр	0,00707
рс	0,00119
ст	0,014432
ту	0,001517
уф	5,22E-05
фх	1E-05
хц	2,61E-05
цч	8,03E-06
чш	0,000151
шщ	0
щъ	3,21E-05
ью	0,000435
юя	9,83E-05

`crossed_bigrams_space(text_with_space)`  
`crossed_bigrams_space(text_without_space)`

аб	0,000787
бв	6,97E-05
вг	0,000202
гд	0,000655
де	0,004144
еэ	1,7E-06
эж	0
жз	0
зи	0,000551
иы	0
ый	0,001765
йк	8,67E-05
кл	0,000498
лм	3,91E-05
мн	0,001576
но	0,010083
оп	0,001417
пр	0,005989
рс	0,000719
ст	0,01211
ту	0,001207
уф	1,36E-05
фх	0
хц	1,7E-06
цч	0
чш	0,000112
шщ	0
щь	2,72E-05
ью	0,000362
юя	1,7E-06
я	0,011411

аб	0,001658
бв	9,43E-05
вг	0,000987
гд	0,000825
де	0,004973
еэ	0,000245
эж	0
жз	6,02E-06
зи	0,000686
иы	0
ый	0,002083
йк	0,000777
кл	0,000688
лм	0,000544
мн	0,002537
но	0,012102
оп	0,003907
пр	0,00707
рс	0,00119
ст	0,014432
ту	0,001517
уф	5,22E-05
фх	1E-05
хц	2,61E-05
цч	8,03E-06
чш	0,000151
шщ	0
щь	3,21E-05
ью	0,000435
юя	9,83E-05
я	0

**Значення  $H$  і  $R$**

	H	R
Letters with spaces	4.391661207126057	0.12166775857478862
Letters without spaces	3.977721444187262	0.19710055981277008
Bigrams with spaces	0.09802607833741339	0.9803947843325174
Bigrams without spaces	0.061206655396903696	0.9876454925153897
Crossed bigrams with spaces	0.19804297423171796	0.19804297423171796
Crossed bigrams without spaces	0.16122355129120827	0.16122355129120827

## CoolPinkProgram



Произвольная часть текста:

он\_делает\_вид\_что\_в\_данном\_конкретном\_случае\_у\_него\_были\_эти\_особые\_причин

Использованные буквы:

й, ф, ц, ы, у, в, к, а, е, п, н, р, г, о, ш, л, щ, д, \_, я, ч, с, м, и,

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:

Поле ввода символов:

Продолжить

Другой

Неравенство для энтропии:

Двоичная таблица угаданных символов:

Вероятности:

Строка состояния:

Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

$q[1] = 0,02$   
 $q[2] = 0$   
 $q[3] = 0$   
 $q[4] = 0,06$   
 $q[5] = 0,02$   
 $q[6] = 0$   
 $q[7] = 0,02$   
 $q[8] = 0,04$   
 $q[9] = 0,04$   
 $q[10] = 0,02$   
 $q[11] = 0$   
 $q[12] = 0,06$   
 $q[13] = 0$   
 $q[14] = 0,06$   
 $q[15] = 0,04$   
 $q[16] = 0,06$   
 $q[17] = 0$   
 $q[18] = 0,02$   
 $q[19] = 0,04$   
 $q[20] = 0,04$   
 $q[21] = 0$   
 $q[22] = 0,02$   
 $q[23] = 0,02$   
 $q[24] = 0,02$   
 $q[25] = 0,12$   
 $q[26] = 0,04$   
 $q[27] = 0,06$   
 $q[28] = 0$   
 $q[29] = 0,02$   
 $q[30] = 0,08$   
 $q[31] = 0,04$   
 $q[32] = 0,04$

[illegible]

