

# КРИПТОГРАФІЯ

## КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ  
джерела відкритого тексту

Виконала: Левашова Світлана

Група: ФБ-13

**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:**

Відсортована таблиця частот символів:

\_: 0.1671

о: 0.0947

а: 0.0726

е: 0.0684

н: 0.0547

т: 0.0533

и: 0.0524

с: 0.0451

л: 0.0410

в: 0.0350

к: 0.0336

р: 0.0328

м: 0.0255

д: 0.0244

у: 0.0243

п: 0.0228

я: 0.0196

ь: 0.0177

з: 0.0143

ы: 0.0141

б: 0.0139

г: 0.0134

ч: 0.0133

й: 0.0084

ж: 0.0084

ш: 0.0080

х: 0.0073

ю: 0.0044

э: 0.0034

ц: 0.0027

щ: 0.0025

ф: 0.0011

Відсортована таблиця частот біграм, що перетинаються (перші 15):

о\_ : 0.0237

е\_ : 0.0200

\_н: 0.0194

а\_ : 0.0178

\_п: 0.0167

\_с: 0.0165

и\_ : 0.0157

\_в: 0.0153

то: 0.0132

я\_ : 0.0132

ь\_ : 0.0116

на: 0.0114

\_т: 0.0104

но: 0.0101

не: 0.0101

Відсортована таблиця частот біграм, що не перетинаються (перші 15):

о\_ : 0.0238

е\_ : 0.0201

\_н: 0.0195

а\_ : 0.0179

\_п: 0.0167

\_с: 0.0165

и\_ : 0.0157

\_в: 0.0152

то: 0.0133

я\_ : 0.0132

ь\_ : 0.0113

на: 0.0112

\_т: 0.0104

не: 0.0103

но: 0.0102

Ентропія H1: 4.357789081405551

$$R = \frac{4,357789081405551}{\log_2(32)(=5)} = 0,8715578162811102$$

Ентропія H2, що перетинаються: 3.9498866184225405

$$R = \frac{3,9498866184225405}{\log_2(32)(=5)} = 0.7899773236845081$$

Ентропія H2, що не перетинаються: 3.9494531633421985

$$R = \frac{3,9494531633421985}{5} = 0.7898906326684397$$

Ентропія H1 для тексту без пробілів: 4.450345774438508

$$R = \frac{4,450345774438508}{5} = 0.8900691548877016$$

Ентропія H2, що перетинаються для тексту без пробілів: 4.132089822886511

$$R = \frac{4,132089822886511}{5} = 0.8264179645773022$$

Ентропія H2, що не перетинаються для тексту без пробілів: 4.131096743499651

$$R = \frac{4,131096743499651}{5} = 0.8262193486999302$$

Произвольная часть текста:  
тавите\_человека\_без\_поддержки\_в\_воздухе\_ч\_него\_будет\_не\_больше\_свободы\_выбо

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: л

Символ по счету: 1

Номер эксперимента: 51

Неравенство для энтропии:  
 $1,99676643671896 < H < 2,80372411601467$

Двоичная таблица угаданных символов:

00001000000000000000000000000000	▲
00000000100000000000000000000000	
<b>00010000000000000000000000000000</b>	
01000000000000000000000000000000	
00000010000000000000000000000000	▼

Поле ввода символов:  
л

Продолжить Другой

Вероятности:

q[1] = 0,4705882
q[2] = 0,1176470
q[3] = 0,0588235
q[4] = 0,0784313
q[5] = 0,0196078
q[6] = 0
q[7] = 0,0392156
q[8] = 0,0392156
q[9] = 0,0196078
q[10] = 0,039215
q[11] = 0
q[12] = 0
q[13] = 0
q[14] = 0,039215
q[15] = 0
q[16] = 0,019607
q[17] = 0
q[18] = 0
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0,019607
q[23] = 0
q[24] = 0
q[25] = 0,019607
q[26] = 0
q[27] = 0
q[28] = 0
q[29] = 0
q[30] = 0
q[31] = 0,019607
q[32] = 0

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

$$1,99676643671896 < H^{(10)} < 2,80372411601467$$

Оцінка надлишковості російської мови:

$$1 - \frac{1,99676643671896}{\log_2(32)(=5)} < R < 1 - \frac{2,80372411601467}{5}; 0,600646713 < R < 0,439255177$$

**Произвольная часть текста:**

авилось\_если\_бы\_кто\_нибудь\_сделал\_то\_же\_самое\_вам\_то\_мое\_место\_я\_его\_первый

---

**Использованные буквы:**

  
  

<b>Порядок n-граммы:</b> 5 символов 10 символов 15 символов <u>20 символов</u> 25 символов 30 символов 35 символов 40 символов 45 символов 50 символов	<b>Введенный символ:</b> _ (пробел)  <b>Символ по счету:</b> 1  <b>Номер эксперимента:</b> 50  <b>Поле ввода символов:</b> <input type="text"/>  <div>Продолжить      Другой</div>	<b>Неравенство для энтропии:</b> $1,50277553998756 < H < 2,27603171251419$  <b>Двоичная таблица угаданных символов:</b> <pre> 1000000000000000000000000000000000 ^ 1000000000000000000000000000000000  1000000000000000000000000000000000   1000000000000000000000000000000000 v 0000000000000100000000000000000000 .....           </pre>
--	---	--

---

**Строка состояния:**

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

$$1,50277553998756 < H^{(20)} < 2,27603171251419$$

### Оцінка надлишковості російської мови:

$$1 - \frac{1,50277553998756}{5} < R < 1 - \frac{2,27603171251419}{5}; 0,699444892 < R < 0,544793657$$

Произвольная часть текста:  
в\_сказать\_я\_здесь\_не\_проповедую\_и\_Богу\_известно\_то\_что\_я\_не\_пытаюсь\_показать

Использованные буквы:  
н.

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: у

Символ по счету: 2

Номер эксперимента: 50

Неравенство для энтропии:  
 $1,63515048660736 < H < 2,25564071234112$

Двоичная таблица угаданных символов:

10000000000000000000000000000000
10000000000000000000000000000000
00000000000100000000000000000000
01000000000000000000000000000000
10000000000000000000000000000000

Вероятности:

q[1] = 0,58
q[2] = 0,18
q[3] = 0
q[4] = 0
q[5] = 0,02
q[6] = 0
q[7] = 0,02
q[8] = 0,02
q[9] = 0
q[10] = 0
q[11] = 0,02
q[12] = 0,02
q[13] = 0
q[14] = 0
q[15] = 0
q[16] = 0,02
q[17] = 0,02
q[18] = 0
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0
q[24] = 0,02
q[25] = 0,02
q[26] = 0
q[27] = 0
q[28] = 0,02
q[29] = 0,02
q[30] = 0
q[31] = 0
q[32] = 0,02

Поле ввода символов:  
у

Продолжить Другой

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

$$1,63515048660736 < H^{(30)} < 2,25564071234112$$

Оцінка надлишковості російської мови:

$$1 - \frac{1,63515048660736}{5} < R < 1 - \frac{2,25564071234112}{5}; 0,672969903 < R < 0,548871858$$

### Висновки:

Реалізовано програму для підрахунку частот букв і біграм в тексті, а також для обчислення значень  $H_1$  та  $H_2$  за безпосереднім означенням.

Виконано підрахунок частот букв, біграм,  $H_1$  та  $H_2$  на російському тексті із заміною імовірностей замінити відповідними частотами. Отримано значення  $H_1$  та  $H_2$  для тексту без пробілів.

Використовуючи програму CoolPinkProgram, оцінено значення ентропії для різних довжин тексту (10, 20, 30).

Робота дозволила освоїти концепції ентропії та надлишковості на символ джерела відкритого тексту.