

# СИМЕТРИЧНА КРИПТОГРАФІЯ КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

## Експериментальна оцінка ентропії на символ джерела відкритого тексту

### 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.

### Постановка задачі:

1. Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму
2. 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. Так одержати значення  $H_1$  та  $H_2$  на тому ж тексті, в якому вилучено всі пробіли.
3. За допомогою програми CoolPinkProgram оцінити значення  $H^{(10)}$ ,  $H^{(20)}$ ,  $H^{(30)}$ .
4. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

### Хід роботи:

1. Написати програму для підрахунку частот букв і частот біграм в тексті роману “Мастер и Маргарита”, підрахувати також  $H_1$  та  $H_2$  на цьому тексті, враховуючи і не враховуючи пробіли.

**!!!УСІ ТАБЛИЦІ ЗНАХОДЯТЬСЯ В ОКРЕМОМУ ФАЙЛІ “charts.pdf”!!!**

*Таблиця частот монограм, враховуючи пробіли:*

$$H_1 = 4.379021219510$$

*Таблиця частот монограм, не враховуючи пробіли:*

$$H_1 = 4.450124790210359$$

*Таблиця частот біграм, що перетинаються, враховуючи пробіли:*

$$H_2 = 7.963835600458476$$

Таблиця частот біграм, що перетинаються, не враховуючи пробіли:

$$H_2 = 8.293071816295$$

Таблиця частот біграм, що не перетинаються, враховуючи пробіли:

$$H_2 = 7.893119625550205$$

Таблиця частот біграм, що не перетинаються, не враховуючи пробіл:

$$H_2 = 8.292505499297492$$

- Провести по 50 експериментів за допомогою програми CoolPinkProgram для знаходження значень  $H^{(10)}$ ,  $H^{(20)}$ ,  $H^{(30)}$ , тобто інтуїтивно вгадувати наступний символ у тексті, якщо відомі відповідно 9, 19 та 29 попередніх.

$$1) \ 1,78803584293028 < H < 2,40085313878716$$

Лабораторная работа №1

Произвольная часть текста:  
а\_не\_совсем\_чистая\_сделка\_о\_которой\_вы\_почти\_забыли\_подвернулась\_вам\_в\_тако

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: е

Символ по счету: 1

Номер эксперимента: 50

Неравенство для энтропии:  
 $1,78803584293028 < H < 2,40085313878716$

Двоичная таблица угаданных символов:

01000000000000000000000000000000
01000000000000000000000000000000
01000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:

q[1] = 0,5
q[2] = 0,22
q[3] = 0
q[4] = 0,04
q[5] = 0
q[6] = 0
q[7] = 0,04
q[8] = 0,04
q[9] = 0,02
q[10] = 0
q[11] = 0,02
q[12] = 0,02
q[13] = 0
q[14] = 0
q[15] = 0
q[16] = 0
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] = 0,04
q[26] = 0,02
q[27] = 0
q[28] = 0
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0

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

$$1) \ 1,18049435650033 < H < 1,93945832095488$$

Лабораторная работа №1

Произвольная часть текста:  
лц\_вы\_попробуете\_нарушить\_обещание\_данное\_ему\_то\_не\_успеете\_вы\_и\_слово\_вымо

Использованные буквы:  
з, с, \_, г, в, д,

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: р

Символ по счету: 7

Номер эксперимента: 50

Неравенство для энтропии:  
 $1,18049435650033 < H < 1,93945832095488$

Двоичная таблица угаданных символов:

10000000000000000000000000000000
10000000000000000000000000000000
01000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:

q[1] = 0,66
q[2] = 0,08
q[3] = 0,04
q[4] = 0,04
q[5] = 0
q[6] = 0,06
q[7] = 0,04
q[8] = 0,02
q[9] = 0,02
q[10] = 0
q[11] = 0
q[12] = 0,02
q[13] = 0
q[14] = 0,02
q[15] = 0
q[16] = 0
q[17] = 0
q[18] = 0
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0
q[24] = 0
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0

Поле ввода символов:  
р

Продолжить Другой

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

2)  $0,997358351226339 < H < 1,57376294916603$

Лабораторная работа №1

Произвольная часть текста:  
ми\_я\_думаю\_шила\_в\_мешке\_не\_утаишь\_и\_что\_бы\_они\_ни\_говорили\_совершенно\_ясно\_

Использованные буквы:  
в, е, о,

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: а

Символ по счету: 4

Номер эксперимента: 50

Неравенство для энтропии:  
 $0,997358351226339 < H < 1,57376294916603$

Двоичная таблица угаданных символов:

00100000000000000000000000000000
0000000000000000000100000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:

q[1] = 0,74
q[2] = 0,06
q[3] = 0,04
q[4] = 0,04
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
q[15] = 0
q[16] = 0,04
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
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0,02
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0

Поле ввода символов:  
а

Продолжить Другой

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

3. Оцінити надлишковість російської мови для кожної з 9-ти моделей джерела.

For monograms  $H_0 = 5.0$

For bigrams  $H_0 = 10.0$

*Monogram entropy considering spaces:*

$$H_1 = 4.379021219510967$$

$$R = 0.12419575609780664$$

*Crossing bigram entropy considering spaces:*

$$H_2 = 3.981917800229238$$

$$R = 0.6018082199770762$$

*No crossing bigram entropy considering spaces:*

$$H_2 = 3.9465598127751025$$

$$R = 0.6053440187224898$$

*Monogram entropy not considering spaces:*

$$H_1 = 4.450124790210359$$

$$R = 0.10997504195792818$$

*Crossing bigram entropy not considering spaces:*

$$H_2 = 4.1465359081475$$

$$R = 0.5853464091852499$$

*No crossing bigram entropy not considering spaces:*

$$H_2 = 4.146252749648746$$

$$R = 0.5853747250351253$$

For  $H^{(10)}$ :

$$1.78803584293028 < H^{(10)} < 2.40085313878716$$
$$0.519829372242568 < R < 0.642392831413944$$

For  $H^{(20)}$ :

$$1.18049435650033 < H^{(20)} < 1.93945832095488$$
$$0.612108335809024 < R < 0.763901128699934$$

For  $H^{(30)}$ :

$$0.997358351226339 < H^{(30)} < 1.57376294916603$$
$$0.685247410166794 < R < 0.8005283297547322$$

### **Висновок:**

Я засвоїла поняття ентропії на символ джерела та його надлишковості, за допомогою деякого тексту, дізналася частоти появлення символів і біграм у російськомовних текстах, знайшла ймовірності відгадування з n-тої спроби символу, якщо відомі попередні символи тексту.