

## 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.

ФБ-11 «Проект Нівечення»

### Хід роботи:

1) Тут має бути таблиця і вона є. Однак я не придумала кращого способу її запису, а те, як вона виглядає наразі, це просто знуцання. З нею, в разі чого, виз можете ознайомитися окремо.

Запишу значення, які ми отримали для  $H_1$  і  $H_{2X}$

#### Для тексту з пробілами:

$H_1$  (Ентропія для одинарних символів): 4.435171349883117

$H_{21}$  (Ентропія для біграм з зсувом 1): 4.017923178097787

$H_{22}$  (Ентропія для біграм з зсувом 2): 4.016769024845158

#### Для тексту без пробілів:

$H_1$  (Ентропія для одинарних символів): 4.473208149868866

$H_{21}$  (Ентропія для біграм з зсувом 1): 4.1268090159753275

$H_{22}$  (Ентропія для біграм з зсувом 2): 4.126810472799593

Код також буде на гітхабі, тож не бачу сенсу дублювати його тут

Перейдемо до проблем, які виникли. Найбільшої насправді стала обробка тексту (видалення пробілів). Як я не намагалася зробити якусь функцію, щоб все обраховувалося в одному скріпті, в мене не вийшло, тому я просто створила два різних з різницею в один символ (я впевнена, що вирішення є, просто я не вважаю його пошук доцільним, так як метою роботи не є навчити мене/нас програмувати (маю сумніви, що це хоч в когось вийде(говорю лише за себе))). Ще проблеми виникли під час обрахунку ентропії для біграм, бо спочатку, ми взагалі не подумали, що там біграми з перетином і без, а потім ще виявилось, що та треба було ділити на два. В принципі більше особливих

проблем не було, якщо не враховувати ті, які виникали, бо я не вмію читати завдання.

## 2. CoolPinkProgram

[illegible]

Лабораторная работа № 1

[illegible]

Произвольная часть текста:  
ебе\_такое\_сильное\_давление\_этого\_закона\_или\_правила\_что\_не\_в\_состоянии\_вынес

Использованные буквы:  
й, ц, у, к, е, 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: г

Символ по счету: 7

Номер эксперимента: 50

Неравенство для энтропии:  
1,74677758254322 < H < 2,35715672729159

Двоичная таблица угаданных символов:  
00000100000000000000000000000000  
00000000000000000000000000000000  
10000000000000000000000000000000  
10000000000000000000000000000000  
10000000000000000000000000000000  
00000000010000000000000000000000

Вероятности:  
q[1] = 0,62  
q[2] = 0,06  
q[3] = 0,02  
q[4] = 0,02  
q[5] = 0  
q[6] = 0,02  
q[7] = 0,02  
q[8] = 0  
q[9] = 0,02  
q[10] = 0  
q[11] = 0,04  
q[12] = 0  
q[13] = 0  
q[14] = 0  
q[15] = 0  
q[16] = 0,02  
q[17] = 0  
q[18] = 0,04  
q[19] = 0  
q[20] = 0  
q[21] = 0  
q[22] = 0  
q[23] = 0,02  
q[24] = 0  
q[25] = 0,04  
q[26] = 0,02  
q[27] = 0,02  
q[28] = 0,02  
q[29] = 0  
q[30] = 0  
q[31] = 0  
q[32] = 0

Поле ввода символов:  
г

Продолжить Другой

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

Проблемою стало те, що я довго не могла знайти екземпляр, бо виявилось, що я забула, що вже знайшла його. Також це було просто банально нудно, мені довелося підключати молодшого брата, але і він довго не витримав.

3.

$$R = 1 - \frac{H_{\infty}}{H_0}$$

Для тексту з першого пункту:

**З пробілами:**

$$R_1 = 0,112966$$

$$R_{21} = 0,196415364$$

$$R_{22} = 0,196646195$$

**Без пробілів:**

$$R_1 = 0,10535837$$

$$R_{21} = 0,174638197$$

$$R_{22} = 0,174637905$$

**Для другого пункту:**

$$R_{10} = 0,549556548$$

$$R_{20} = 0,630289658$$

$$R_{30} = 0,650644483$$

**Висновки:** Під час виконання роботи дізналися про поняття ентропії та надлишковості мови, навчилися їх визначати та оцінювати. Нами було експериментально підтверджено, що найбільш уживаним символом є пробіл (для тексту з пробілами) або літера о (для тексту без пробілів). Також експериментальним шляхом підтверджено, що зі збільшенням значення ентропії (на практиці зі збільшенням символів, які нам достеменно відомі) збільшується значення  $R$  (тобто, ми підтвердили достатньо тривіальне твердження, що чим більше у нас початкової інформації, ти легше вгадати наступну літеру). Також ми побачили, що значення надлишковості може суттєво коливатися (у проміжку між 0 та 1) для однієї й тої самої мови.