

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

СИМЕТРИЧНА КРИПТОГРАФІЯ
КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконала:
студентка групи ФІ-93
Меднікова Олександра

Перевірив:
Чорний О. М.

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.

Постановка задачі

1. 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, де імовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 на тому ж тексті, в якому видалено всі пробіли.
2. За допомогою CoolPinkProgram оцінити значення $H^{(10)}$, $H^{(20)}$, $H^{(30)}$.
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи

1. Програмна реалізація на github: https://github.com/OleksandraMednikova/fi-labs-2022/tree/master/cp_1/mednikova_fi-93_cp1
2. Частоти букв та біграм(з перетином та без) з пробілом та без у алфавіті наведені у файлі result.txt

Значення H_1 та H_2 :

H_1 у тексті з пробілом: 4.383831209580581

H_2 біграм без перетину у тексті з пробілом: 3.985189847546685

H_2 біграм з перетином у тексті з пробілом: 3.9856911068555023

H_1 у тексті без пробілу: 4.46869156754106

H_2 біграм без перетину у тексті без пробілу: 4.1520236191862425

H_2 біграм з перетином у тексті без пробілу: 4.152704651455609

3. За допомогою CoolPinkProgram були отримані такі значення:

$$1,6563 \leq H^{(10)} \leq 2,3195$$

The screenshot displays the CoolPinkProgram interface with the following elements:

- Произвольная часть текста:** тельствах_или_что_случилось_нечто_непредвиденное_освобождающее_его_от_необх
- Использованные буквы:** (empty field)
- Порядок n-граммы:** A list with options: 5 символов, 10 символов (selected), 15 символов, 20 символов, 25 символов, 30 символов, 35 символов, 40 символов, 45 символов, 50 символов.
- Введенный символ:** _ (пробел)
- Символ по счету:** 1
- Номер эксперимента:** 50
- Неравенство для энтропии:** $1,65632928479958 < H < 2,31954462393373$
- Двоичная таблица угаданных символов:** A table with 32 rows and 2 columns of 0s and 1s.
- Вероятности:** A list of probabilities $q[1]$ through $q[32]$ ranging from 0 to 0.56.
- Поле ввода символов:** (empty field)
- Buttons:** Продолжить, Другой
- Строка состояния:** Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

$$1,4902 \leq H^{(20)} \leq 2,2201$$

Лабораторная работа №1

Произвольная часть текста:
илось_если_бы_кто_нибудь_сделал_то_же_самое_вам_то_мое_место_я_его_первый_з

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: и

Символ по счету: 1

Номер эксперимента: 50

Неравенство для энтропии:
 $1,49017571707505 < H < 2,22006166784613$

Двоичная таблица угаданных символов:

10000000000000000000000000000000
00000000000000000000000001000000000
001000000000000000000000000000000
00000000000000000000010000000000000
100000000000000000000000000000000

Поле ввода символов:
и

Продолжить Другой

Вероятности:

q[1] = 0,62
q[2] = 0,1
q[3] = 0,04
q[4] = 0,04
q[5] = 0
q[6] = 0
q[7] = 0,04
q[8] = 0,02
q[9] = 0
q[10] = 0,02
q[11] = 0
q[12] = 0
q[13] = 0,02
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
q[22] = 0
q[23] = 0,02
q[24] = 0,02
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0,02
q[29] = 0
q[30] = 0
q[31] = 0,02
q[32] = 0

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

$$2,2602 \leq H^{(30)} \leq 2,7294$$

Произвольная часть текста:
ловека_я_бы_не_удивлялся_да_и_кто_я_такой_в_конце_концов_я_сам_такой_же_то_

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: _ (пробел)

Символ по счету: 1

Номер эксперимента: 50

Неравенство для энтропии:
 $2,26015370336669 < H < 2,72943203625482$

Двоичная таблица угаданных символов:

10000000000000000000000000000000
0000000000000000000001000000000000
00000000000000000000000000000000
00000000000000000000000000000000
00000000000000000000000000000000

Поле ввода символов:

Продолжить Другой

Вероятности:

q[1] = 0,52
q[2] = 0,08
q[3] = 0,06
q[4] = 0,02
q[5] = 0
q[6] = 0
q[7] = 0
q[8] = 0
q[9] = 0
q[10] = 0
q[11] = 0
q[12] = 0
q[13] = 0
q[14] = 0,02
q[15] = 0
q[16] = 0,02
q[17] = 0,04
q[18] = 0,02
q[19] = 0,04
q[20] = 0,02
q[21] = 0
q[22] = 0,02
q[23] = 0
q[24] = 0
q[25] = 0,06
q[26] = 0
q[27] = 0,02
q[28] = 0
q[29] = 0,02
q[30] = 0
q[31] = 0
q[32] = 0,04

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

4. Оцінка надлишковості

$$H_0 = \log_2 32 = 5 \text{ (алфавіт з пробілом)}$$

$$H_{inf} = H^{(10)}, \text{ тоді } 1 - \frac{1,6563}{5} \leq R \leq 1 - \frac{2,3195}{5}, \text{ тобто } 0,536 \leq R \leq 0,669$$

$$H_{inf} = H^{(20)}, \text{ тоді } 1 - \frac{1,4902}{5} \leq R \leq 1 - \frac{2,2201}{5}, \text{ тобто } 0,556 \leq R \leq 0,702$$

$$H_{inf} = H^{(30)}, \text{ тоді } 1 - \frac{2,2602}{5} \leq R \leq 1 - \frac{2,7294}{5}, \text{ тобто } 0,454 \leq R \leq 0,548$$

$$H_{inf} = H_1, \text{ тоді } R = 1 - \frac{4,3838}{5} = 0,12324$$

$$H_{inf} = H_{21}, \text{ тоді } R = 1 - \frac{3,9852}{5} = 0,2029$$

$$H_{inf} = H_{22}, \text{ тоді } R = 1 - \frac{3,9856}{5} = 0,2028$$

$$H_0 = \log_2 31 = 4,9542 \text{ (алфавіт без пробілу)}$$

$$H_{inf} = H_1, \text{ тоді } R = 1 - \frac{4,4686}{4,9542} = 0,100$$

$$H_{inf} = H_{21}, \text{ тоді } R = 1 - \frac{4,1520}{4,9542} = 0,162$$

$$H_{inf} = H_{22}, \text{ тоді } R = 1 - \frac{4,1527}{4,9542} = 0,162$$