



Міністерство освіти і науки, молоді та спорту України
Національний технічний університет України
"Київський політехнічний інститут"
Фізико-Технічний інститут

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

за семестровий курс предмету

«Симетрична криптографія»

Роботу виконали:

Студент групи ФІ-03
Гілевський Олександр

Приймав:

Чорний Олег Миколайович

Київ-2023

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.

Порядок виконання роботи:

0. Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.
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 (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 на тому ж тексті, в якому вилучено всі пробіли.
2. За допомогою програми CoolPinkProgram оцінити значення $H(10)$, $H(20)$, $H(30)$.
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи

- Очистка вхідного текстового файлу та редагування;
- Написання функцій для підрахунку частот літер та частот біграм. Визначили частоти та значення ентропії;
- Обрахували оцінки надлишковості російської мови для даних умовних ентропій джерел;
- Оформлення звіту.

1. Результати частот букв, та H_1 і H_2 :

З перетином, але без пробілів

Буква	Кількість
о	95273
е	73163
а	65926
т	53931
и	53630
н	53037
с	43046
л	38362
в	37597
р	33528
к	26994
м	25946
д	25711
у	24780
п	21133
ь	19375
я	17406
ч	14872
б	14412
г	14001
ы	13920
з	12470
ж	9574
й	8441
х	7068
ш	6911
ю	4730
э	2522
щ	2520
ц	2319

ф

972

Н1: 4,444373807624386

Н2: 4,130122805459284

Без перетину та пробілів

Буква	Кількість
о	95273
е	73163
а	65926
т	53931
и	53630
н	53037
с	43046
л	38362
в	37597
р	33528
к	26994
м	25946
д	25711
у	24780
п	21133
ь	19375
я	17406
ч	14872
б	14412
г	14001
ы	13920
з	12470
ж	9574
й	8441

х	7068
ш	6911
ю	4730
э	2522
щ	2520
ц	2319
ф	972

Н1: 4,444373807624386

Н2: 4,130127843664438

Без перетину, але з пробілами

Символ	Кількість
" "	167105
о	95273
е	73163
а	65926
т	53931
и	53630
н	53037
с	43046
л	38362
в	37597
р	33528
к	26994
м	25946
д	25711
у	24780
п	21133
ь	19375
я	17406

ч	14872
б	14412
г	14001
ы	13920
з	12470
ж	9574
й	8441
х	7068
ш	6911
ю	4730
э	2522
щ	2520
ц	2319
ф	972

H1: 4,349377539781125

H2: 3,9499937912023224

З перетином і пробілами

Символ	Кількість
" "	167105
о	95273
е	73163
а	65926
т	53931
и	53630
н	53037
с	43046
л	38362
в	37597
р	33528

к	26994
м	25946
д	25711
у	24780
п	21133
ь	19375
я	17406
ч	14872
б	14412
г	14001
ы	13920
з	12470
ж	9574
й	8441
х	7068
ш	6911
ю	4730
э	2522
щ	2520
ц	2319
ф	972

H1: 4,349377539781125

H2: 3,9501128865434714

Результати біграм в файлі **results.txt**

2. Оцінки для $H^{(10)}$, $H^{(20)}$, $H^{(30)}$

- $H^{(10)}$

Лабораторная работа №1

Произвольная часть текста:
сб_что_вы

Использованные буквы:

Порядок n-граммы:
5 символов

Введенный символ:

Символ по счету:

Номер эксперимента: 51

Неравенство для энтропии:
 $2.46549310415541 < H < 3.1164130274445$

Двоичная таблица угаданных символов:

Вероятности:

q[1] = 0.44
q[2] = 0.12
q[3] = 0.02
q[4] = 0.02
q[5] = 0
q[6] = 0.02
q[7] = 0.02
q[8] = 0.02
q[9] = 0.02
q[10] = 0.04
q[11] = 0.02
q[12] = 0
q[13] = 0.02
q[14] = 0
q[15] = 0
q[16] = 0.06
q[17] = 0.04
q[18] = 0.04
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0.02
q[23] = 0
q[24] = 0
q[25] = 0.02
q[26] = 0
q[27] = 0.02
q[28] = 0
q[29] = 0
q[30] = 0.04
q[31] = 0
q[32] = 0

Строка состояния:

$$R = 1 - \frac{2.4655 + 3.1164}{2} \cdot \frac{1}{\log_2(32)} = 0.442$$

- $H^{(20)}$

Лабораторная работа №1

Произвольная часть текста:
азать_что_то_что_он

Использованные буквы:

Порядок n-граммы:
10 символов

Введенный символ:

Символ по счету:

Номер эксперимента: 51

Неравенство для энтропии:
 $2.56492386160696 < H < 3.13669456263237$

Двоичная таблица угаданных символов:

Вероятности:

q[1] = 0.42
q[2] = 0.06
q[3] = 0.1
q[4] = 0.04
q[5] = 0
q[6] = 0.06
q[7] = 0.02
q[8] = 0
q[9] = 0
q[10] = 0.02
q[11] = 0
q[12] = 0.02
q[13] = 0
q[14] = 0
q[15] = 0.02
q[16] = 0.02
q[17] = 0.04
q[18] = 0
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0.08
q[24] = 0.02
q[25] = 0.02
q[26] = 0
q[27] = 0
q[28] = 0
q[29] = 0.02
q[30] = 0.02
q[31] = 0
q[32] = 0.02

Строка состояния:

$$R = 1 - \frac{2.5649 + 3.1367}{2} \cdot \frac{1}{\log_2(32)} = 0.43$$

- $H^{(30)}$

Лабораторная работа №1

Произвольная часть текста:
_ между_справедливости_и_неспра

Использованные буквы:

Порядок n-граммы:
5 символов
10 символов
15 символов
20 символов
25 символов
35 символов
40 символов
45 символов
50 символов

Введенный символ:

Символ по счету:

Номер эксперимента: 51

Неравенство для энтропии:
 $2.59769355730614 < H < 3.20888737212111$

Двоичная таблица угаданных символов:

Поле ввода символов:

Продолжить Другой

Вероятности:

q[1] = 0.42
q[2] = 0.08
q[3] = 0.08
q[4] = 0.02
q[5] = 0
q[6] = 0.02
q[7] = 0.02
q[8] = 0.02
q[9] = 0.02
q[10] = 0.02
q[11] = 0.02
q[12] = 0
q[13] = 0
q[14] = 0.06
q[15] = 0.02
q[16] = 0
q[17] = 0
q[18] = 0
q[19] = 0
q[20] = 0
q[21] = 0.04
q[22] = 0.02
q[23] = 0.04
q[24] = 0
q[25] = 0
q[26] = 0.06
q[27] = 0
q[28] = 0
q[29] = 0.02
q[30] = 0.02
q[31] = 0
q[32] = 0

Строка состояния:

$$R = 1 - \frac{2.5977 + 3.2089}{2} \cdot \frac{1}{\log_2(32)} = 0.42$$

Висновок: в цій лабі розглянули як рахувати частоту літер в тексті, біграм, ентропію та надлишковість. Перегналися, що найчастішим символом є пробіл, а після нього голосні літери (а, о і тд).