

Міністерство освіти і науки України
Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського"
Фізико-технічний інститут

СИМЕТРИЧНА КРИПТОГРАФІЯ
КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконав: студент гр. ФІ-94, Кріпака І.А.

Перевірив: Чорний О.М.

Київ – 2021

1) Мета комп'ютерного практикуму

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

2) Постановка задачі

Створити програму для експериментальної оцінки ентропії на символ джерела відкритого тексту, порівняти різні моделі джерела відкритого тексту для наближеного визначення ентропії.

3) 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

4) Опис труднощів

Реалізуючи програмний код у мене виникла проблема із чищенням файлу від інших "непотрібних" символів, вирішилася досить легко за допомогою регулярних виразів.

5) Практична частина

[Посилання на код](#)

```
Total letters sum: 5696814.0
Total letters sum: 4793739.0
Total letters sum: 5681234.0
Total letters sum: 4793738.0
Entropy for:
One symbol with gap (H1): 4.3840365490920545
One symbol without gaps (H1): 4.460337963530066
Bigram with gaps (H2): 4.000823175252397
Bigram without gaps (H2): 4.153176280546273
```

6) Результати підрахунку частот у файлі cp1_probability_tables.ods

7) Значения энтропии

	Алфавит із пробілом	Алфавит без пробілу
H0	5	4.954196
H1	4.3840365490920545	4.460337963530066
H2	4.000823175252397	4.153176280546273

Надлишковість для H(10)

$$2.43069086962413 < H(10) < 2.99532746619176$$

$$H(10) = 2.9, H_0 = 5$$

$$R = 1 - 2.9/5 = 1 - 0.58 = \mathbf{0.42}$$

Лабораторная работа №1

Произвольная часть текста:

ь_настолько_глубоко_мы_испытываем_на_себе_такое_сильное_давление_этого_закон

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: **к**

Символ по счету: **1**

Номер эксперимента: **66**

Поле ввода символов:

к

Продолжить

Другой

Неравенство для энтропии:
2,43069086962413 < H < 2,99532746619176

Двоичная таблица угаданных символов:

00000000001000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
01000000000000000000000000000000

Вероятности:

q[1] = 0,5151515!
q[2] = 0,0454545!
q[3] = 0,0303030!
q[4] = 0,0151515!
q[5] = 0
q[6] = 0,0151515!
q[7] = 0
q[8] = 0
q[9] = 0
q[10] = 0
q[11] = 0,0454545!
q[12] = 0,0303030!
q[13] = 0,0151515!
q[14] = 0,0303030!
q[15] = 0
q[16] = 0,0151515!
q[17] = 0,0151515!
q[18] = 0,0303030!
q[19] = 0
q[20] = 0,0303030!
q[21] = 0
q[22] = 0,0151515!
q[23] = 0,0151515!
q[24] = 0,0151515!
q[25] = 0
q[26] = 0
q[27] = 0,0303030!
q[28] = 0,0151515!
q[29] = 0
q[30] = 0
q[31] = 0,0303030!
q[32] = 0,0454545!

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

Надлишковість для H(20)

$$2.8193332347205 < H(20) < 3.5677693577254$$

$$H(20) = 3.2, H_0 = 5$$

$$R(20) = 1 - 3.2/5 = 1 - 0.64 = \mathbf{0.36}$$

Лабораторная работа №1

Произвольная часть текста:
мер_вы_страшно_устали_когда_были_так_несправедливы_к_детям_та_не_совсем_чис

Использованные буквы:
в, и, т, _

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: 5

Номер эксперимента: 65

Неравенство для энтропии:
 $2,8193332347205 < H < 3,56776973577254$

Двоичная таблица угаданных символов:

10000000000000000000000000000000
01000000000000000000000000000000
00000000000000000000000000000000
00000000000000000000000000000000
00000000000000000000000000000000

Поле ввода символов:
л

Продолжить Другой

Вероятности:

q[1] = 0,36923071
q[2] = 0,12307692
q[3] = 0,0307692
q[4] = 0,0153846
q[5] = 0,0153846
q[6] = 0,0153846
q[7] = 0,0153846
q[8] = 0,0461538
q[9] = 0,0307692
q[10] = 0,0153846
q[11] = 0
q[12] = 0,0307692
q[13] = 0
q[14] = 0,0307692
q[15] = 0
q[16] = 0,0153846
q[17] = 0,0461538
q[18] = 0
q[19] = 0,0307692
q[20] = 0,0307692
q[21] = 0
q[22] = 0
q[23] = 0,0153846
q[24] = 0,0153846
q[25] = 0,0153846
q[26] = 0,0307692
q[27] = 0,0153846
q[28] = 0,0307692
q[29] = 0,0153846
q[30] = 0
q[31] = 0
q[32] = 0

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

Надлишковість для $H(30)$
 $2.54709563560827 < H(30) < 3.32833585338233$
 $H(30) = 3, H_0 = 5$
 $R(30) = 1 - 3/5 = 1 - 0.6 = 0.4$

Лабораторная работа №1

Произвольная часть текста:
порядочность_настолько_глубоко_мы_испытываем_на_себе_такое_сильное_давление

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: о

Символ по счету: 1

Номер эксперимента: 61

Неравенство для энтропии:
 $2,54709563560827 < H < 3,32833585338233$

Двоичная таблица угаданных символов:

10000000000000000000000000000000
10000000000000000000000000000000
00010000000000000000000000000000
00000000000000000000000000000000
10000000000000000000000000000000

Поле ввода символов:
о

Продолжить Другой

Вероятности:

q[1] = 0,44262291
q[2] = 0,0819672
q[3] = 0,0327868
q[4] = 0,0327868
q[5] = 0,0327868
q[6] = 0,0327868
q[7] = 0
q[8] = 0,0163934
q[9] = 0
q[10] = 0,0327868
q[11] = 0,0163934
q[12] = 0
q[13] = 0,0163934
q[14] = 0,0163934
q[15] = 0
q[16] = 0
q[17] = 0,0327868
q[18] = 0,0163934
q[19] = 0,0163934
q[20] = 0
q[21] = 0,0163934
q[22] = 0,0163934
q[23] = 0
q[24] = 0,0163934
q[25] = 0,0163934
q[26] = 0,049180
q[27] = 0,0163934
q[28] = 0
q[29] = 0
q[30] = 0,0327868
q[31] = 0
q[32] = 0,0163934

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

	Для $H(10)$	Для $H(20)$	Для $H(30)$
$R(*)$	0.42	0.36	0.4