



НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

Кафедра Інформаційної Безпеки

Лабораторна робота №1 дисципліни

”КРИПТОГРАФІЯ”

Підготував:

студент групи ФБ-06

Жак Костянтин

Київ 2022

Тема роботи: 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 \square_1 та \square_2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення \square_1 та \square_2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення \square_1 та \square_2 на тому ж тексті, в якому вилучено всі пробіли.
2. За допомогою програми CoolPinkProgram оцінити значення $\square^{(10)}$, $\square^{(20)}$, $\square^{(30)}$.
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи

За початковий текст взяв “Хроніки нарнії”

Значення частот букв з пробілами:

Буква	Її частотність
а	0.06865
б	0.01302
в	0.03577
г	0.0144
д	0.0275
е	0.0623
ж	0.00803
з	0.01467
и	0.05746
й	0.00799
к	0.02897
л	0.04541
м	0.02643
н	0.05237
о	0.09153
п	0.02322
р	0.03641
с	0.04534
т	0.04996
у	0.02378
ф	0.00066
х	0.00749
ц	0.00244
ч	0.01247
ш	0.00784
щ	0.00225
ы	0.01605
ь	0.01802
э	0.00314
ю	0.0049
я	0.0159
пробел	0.17026

Ентропія:

4.354901763494896

Надлишковість:

0.1290196473010209

Значення частот букв без пробілів:

Буква	Її частотність
а	0.08274
б	0.01569
в	0.04312
г	0.01735
д	0.03315
е	0.07508
ж	0.00968
з	0.01768
и	0.06925
й	0.00963
к	0.03491
л	0.05473
м	0.03185
н	0.06312
о	0.11031
п	0.02799
р	0.04388
с	0.05465
т	0.06021
у	0.02866
ф	0.0008
х	0.00903
ц	0.00294
ч	0.01503
ш	0.00944
щ	0.00271
ы	0.01935
ь	0.02172
э	0.00378
ю	0.0059
я	0.01917

Ентропія:

4.45752152505319

Надлишковість:

0.10025335174796479

Значення частот біграм що перетинаються з пробілами:

Значення частот біграм що не переитнаються з пробілами:

1. $\square(10)$

Лабораторная работа №1

Произвольная часть текста:
их_как_ошибаются_скажем_при_сложении_чисел_но_понятие_о_добре_и_зле_не_в_б

Использованные буквы:
н,_,д,м,с,т,у.

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: ш

Символ по счету: 8

Номер эксперимента: 50

Неравенство для энтропии:
 $2.26476438758702 < H < 2.97669456263237$

Двоичная таблица угаданных символов:

00100000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:

q[1] = 0.42
q[2] = 0.12
q[3] = 0.04
q[4] = 0.1
q[5] = 0.08
q[6] = 0
q[7] = 0.02
q[8] = 0.02
q[9] = 0.04
q[10] = 0.02
q[11] = 0
q[12] = 0
q[13] = 0
q[14] = 0
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] = 0.02
q[24] = 0.02
q[25] = 0.02
q[26] = 0.02
q[27] = 0.02
q[28] = 0
q[29] = 0.04
q[30] = 0
q[31] = 0
q[32] = 0

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

$$2.265 < \square^{(10)} < 2.980$$

$$2. \square^{(20)}$$

Лабораторная работа №1

Произвольная часть текста:
живет_вокруг_или_во

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:

Символ по счету:

Номер эксперимента: 51

Неравенство для энтропии:
 $2.08036693483386 < H < 2.93585768303304$

Двоичная таблица угаданных символов:

10000000000000000000000000000000
10000000000000000000000000000000
00100000000000000000000000000000
00010000000000000000000000000000
00000100000000000000000000000000

Вероятности:

q[1] = 0.42
q[2] = 0.14
q[3] = 0.12
q[4] = 0.04
q[5] = 0.04
q[6] = 0.02
q[7] = 0.02
q[8] = 0.04
q[9] = 0.02
q[10] = 0
q[11] = 0
q[12] = 0.02
q[13] = 0.02
q[14] = 0
q[15] = 0.02
q[16] = 0
q[17] = 0.02
q[18] = 0
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0.04
q[23] = 0.02
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

Строка состояния:

$$2.080 < \square^{(20)} < 2.940$$

$$3. \square^{(30)}$$

Произвольная часть текста:
ственный_потому_что_люди_дума

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:

Символ по счету:

Номер эксперимента: 51

Неравенство для энтропии:
 $1.45140622963482 < H < 2.18205897724832$

Двоичная таблица угаданных символов:

10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
00100000000000000000000000000000
10000000000000000000000000000000

Вероятности:

q[1] = 0.62
q[2] = 0.08
q[3] = 0.06
q[4] = 0.02
q[5] = 0.04
q[6] = 0
q[7] = 0.02
q[8] = 0.02
q[9] = 0.06
q[10] = 0
q[11] = 0
q[12] = 0
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
q[22] = 0.02
q[23] = 0
q[24] = 0.02
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0
q[29] = 0
q[30] = 0.02
q[31] = 0
q[32] = 0

Строка состояния:

$$1.451 < H^{(30)} < 2.182$$

3. Оцінимо надлишковість на основі отриманих результатів:

Для нижньої границі:

- 1) $R1 = 1 - \frac{2.26476438758702}{5} = 0.547047122482596$
- 2) $R2 = 0.583926613033228$
- 3) $R3 = 0.709718754073036$

Для верхньої границі:

- 1) $R1 = 0.404661087473526$
- 2) $R2 = 0.412828463393392$
- 3) $R3 = 0.563588204550336$

Висновок

Під час виконання даного лабораторного практикуму було написано програму, що обраховує частотність літер та біграм російської мови на основі тексту “Хроніки Нарнії”, на основі цих результатів було обраховано надлишковість російської мови. У шифруванні слід використовувати текст без пробілів, так як ентропія в ньому більша.

Також навчився працювати з програмою CoolPinkProgram.