

СИМЕТРИЧНА КРИПТОГРАФІЯ

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконали студенти групи ФІ-94
Костюк Кирило і Панасюк Єгор
Варіант-4

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.

Хід роботи

Написали програму для підрахунку частот букв і частот біграм в тексті, а також підрахунку H_1 та H_2 . Підраховували частоти букв і біграм, а також значення H_1 та H_2 на тексті Гоголя “Мертві душі”. За допомогою програми CoolPinkProgram оцінили значення $H(10)$, $H(20)$, $H(30)$

Output

Монограми з пробілом	Монограми без пробілу
letter: frequency: 0.158805	letter: a frequency: 0.0767927
letter: а frequency: 0.0645977	letter: б frequency: 0.0191969
letter: б frequency: 0.0161484	letter: в frequency: 0.0467936
letter: в frequency: 0.0393625	letter: г frequency: 0.0175203
letter: г frequency: 0.014738	letter: д frequency: 0.031274
letter: д frequency: 0.0263075	letter: е frequency: 0.0855975
letter: е frequency: 0.0720042	letter: ж frequency: 0.0111428
letter: ж frequency: 0.00937329	letter: з frequency: 0.0168903
letter: з frequency: 0.0142081	letter: и frequency: 0.0683831
letter: и frequency: 0.0575235	letter: й frequency: 0.010541
letter: й frequency: 0.008867	letter: к frequency: 0.0413734
letter: к frequency: 0.0348031	letter: л frequency: 0.0447019
letter: л frequency: 0.037603	letter: м frequency: 0.0297693
letter: м frequency: 0.0250418	letter: н frequency: 0.061827

letter: н frequency: 0.0520086	letter: о frequency: 0.112799
letter: о frequency: 0.0948859	letter: п frequency: 0.0278579
letter: п frequency: 0.0234339	letter: р frequency: 0.0413503
letter: р frequency: 0.0347837	letter: с frequency: 0.0510562
letter: с frequency: 0.0429483	letter: т frequency: 0.0616898
letter: т frequency: 0.0518932	letter: у frequency: 0.0303116
letter: у frequency: 0.025498	letter: ф frequency: 0.00120043
letter: ф frequency: 0.00100979	letter: х frequency: 0.00975555
letter: х frequency: 0.00820632	letter: ц frequency: 0.00316807
letter: ц frequency: 0.00266497	letter: ч frequency: 0.0190035
letter: ч frequency: 0.0159856	letter: ш frequency: 0.00957201
letter: ш frequency: 0.00805193	letter: щ frequency: 0.00296304
letter: щ frequency: 0.0024925	letter: ы frequency: 0.018901
letter: ы frequency: 0.0158994	letter: ь frequency: 0.0212869
letter: ь frequency: 0.0179065	letter: э frequency: 0.00225039
letter: э frequency: 0.00189302	letter: ю frequency: 0.0060038
letter: ю frequency: 0.00505036	letter: я frequency: 0.0190266
letter: я frequency: 0.0160051	

Частоти біграм можна подивитися за даним посиланням:

<https://docs.google.com/document/d/1oz12Z8hmy2ekJH4YoBJ72HYR3jsBvWS58CaBcTp2UJk/edit?usp=sharing>

H0=5	З пробілом	Без пробілу
H1	4.38921	4.46717
H2(перетин)	3.97137	4.13174
H2(не перетин)	3.97131	4.13155

З програмки:

Лабораторная работа №1

Произвольная часть текста:
ими_эле_не_более_чем_в_цвете_их_волос_я_знаю_что_по_мнению_некоторых_людей_

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: e

Символ по счету: 1

Номер эксперимента: 52

Неравенство для энтропии:
 $1.86328911854908 < H < 2.58769131187048$

Двоичная таблица угаданных символов:

00100000000000000000000000000000	▲
00000000000001000000000000000000	
00010000000000000000000000000000	
10000000000000000000000000000000	
00010000000000000000000000000000	▼

Поле ввода символов:
e

Продолжить Другой

Вероятности:

q[1] = 0.5
q[2] = 0.1346153
q[3] = 0.0769230
q[4] = 0.0769230
q[5] = 0
q[6] = 0.0192307
q[7] = 0.0192307
q[8] = 0
q[9] = 0
q[10] = 0
q[11] = 0.019230
q[12] = 0
q[13] = 0
q[14] = 0.019230
q[15] = 0
q[16] = 0.019230
q[17] = 0
q[18] = 0
q[19] = 0.038461
q[20] = 0
q[21] = 0.019230
q[22] = 0.019230
q[23] = 0
q[24] = 0
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0.038461
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0

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

$$1.86 < H(10) < 2.58$$

Лабораторная работа №1

Произвольная часть текста:
ей_семьи_или_к_тем_кто_живет_вокруг_или_вообще_ко_всем_людям_однако_они_все

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: 52

Неравенство для энтропии:
 $1.51357636042729 < H < 2.29214029359501$

Двоичная таблица угаданных символов:

10000000000000000000000000000000	▲
00010000000000000000000000000000	
00000000000000010000000000000000	
10000000000000000000000000000000	
10000000000000000000000000000000	▼

Поле ввода символов:
к

Продолжить Другой

Вероятности:

q[1] = 0.5961538
q[2] = 0.1346153
q[3] = 0.0384615
q[4] = 0.0192307
q[5] = 0.0192307
q[6] = 0.0192307
q[7] = 0
q[8] = 0
q[9] = 0
q[10] = 0.019230
q[11] = 0.019230
q[12] = 0.019230
q[13] = 0
q[14] = 0
q[15] = 0.019230
q[16] = 0.038461
q[17] = 0.019230
q[18] = 0
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0
q[24] = 0.019230
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0
q[29] = 0
q[30] = 0
q[31] = 0.019230
q[32] = 0

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

$$1.51 < H(20) < 2.29$$

Лабораторная работа №1

Произвольная часть текста:
ледует_иметь_одну_или_четырёх_но_они_всегда_были_согласны_в_том_что_брать_к

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: 52

Неравенство для энтропии:
 $1.34798667009934 < H < 1.94393463008125$

Двоичная таблица угаданных символов:

00000000000000000000000000000000	00000000000000000000000000000000
10000000000000000000000000000000	00000000000000000000000000000000
00100000000000000000000000000000	00000000000000000000000000000000
10000000000000000000000000000000	00000000000000000000000000000000
00000000000000000000000000000000	00000000000000000000000000000000

Вероятности:

q[1] = 0.6538461
q[2] = 0.0961538
q[3] = 0.0192307
q[4] = 0
q[5] = 0
q[6] = 0.0384615
q[7] = 0
q[8] = 0
q[9] = 0.0576923
q[10] = 0
q[11] = 0.038461
q[12] = 0.038461
q[13] = 0.019230
q[14] = 0
q[15] = 0
q[16] = 0.019230
q[17] = 0
q[18] = 0
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0.019230
q[23] = 0
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

Поле ввода символов:

Продолжить Другой

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

$$1.34 < H(30) < 2.94$$

Всі результати з програмки:

$$1.86 < H(10) < 2.58$$

$$1.51 < H(20) < 2.29$$

$$1.34 < H(30) < 2.94$$

Оцінка надлишковості

При $H_1 = 4.38921$ (з пробілом):

$$R = 1 - 0.877842 = 0.122158$$

При $H_1 = 4.467171$ (без пробілу):

$$R = 1 - 0.8934342 = 0.1065658$$

При $H_2 = 3.97137$ (перетин з пробілом):

$$R = 1 - 0.794274 = 0.205726$$

При $H_2 = 4.13174$ (перетин без пробілу):

$$R = 1 - 0.826348 = 0.173652$$

При $H_2 = 3.97131$ (не перетин з пробілом):

$$R = 1 - 0.794262 = 0.205738$$

При $H_2 = 4.13155$ (не перетин без пробілу):

$$R = 1 - 0.82631 = 0.17369$$

Нерівності з даних з програмки:

При $H(10)$:

$$0.484 < R < 0.628$$

При $H(20)$:

$$0.542 < R < 0.698$$

При $H(30)$:

$$0.412 < R < 0.732$$

Висновок

Ми набули практичних навичок, щодо оцінки ентропії на символ джерела. Написали програмний засіб, який допомагає в цьому. Лаба весела, але наступні веселіші.