

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

КРИПТОГРАФІЯ
КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконав: ФБ-11 Тимощук Ілья

Київ – 2023

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

Для кращої демонстрації пробіл « » було замінено на нижнє підкреслення «_». Відредаговані тексти з пробілом та без були збережені в text_edited_w_space.txt та text_edited_no_space.txt відповідно.

Таблиця частот символів з пробілом

Symbol	Frequency
	0.15429578
о	0.09884576
е	0.07167816
и	0.06160380
а	0.06058887
н	0.05544288
т	0.05256577
с	0.04503854
л	0.03911842
в	0.03778243
р	0.03673361
м	0.03082955
к	0.02703382
д	0.02587798
п	0.02374646
у	0.02193243
я	0.02125105
ы	0.01643505
ь	0.01540229
г	0.01537375
б	0.01479761
з	0.01337957
ч	0.01274992
х	0.00931986
й	0.00880258
ж	0.00784473
ш	0.00652301
ю	0.00614665
щ	0.00299306
ц	0.00240800
э	0.00227066
ф	0.00097212
ъ	0.00021404
ё	0.00000178

Таблиця частот символів без пробіла

Symbol	Frequency
о	0.11687983
е	0.08475559
и	0.07284320
а	0.07164310
н	0.06555825
т	0.06215621
с	0.05325566
л	0.04625544
в	0.04467570
р	0.04343553
м	0.03645429
к	0.03196605
д	0.03059933
п	0.02807892
у	0.02593393
я	0.02512824
ы	0.01943357
ь	0.01821238
г	0.01817864
б	0.01749738
з	0.01582062
ч	0.01507610
х	0.01102023
й	0.01040858
ж	0.00927598
ш	0.00771311
ю	0.00726808
щ	0.00353913
ц	0.00284733
э	0.00268493
ф	0.00114948
ъ	0.00025310
ё	0.00000211

Оскільки таблиці частот біграм доволі великі, тут будуть вказані лише перші 10 частот з кожної таблиці. Повна версія цих та інших таблиць доступна у відповідних excel файлах.

Таблиця частот біграм_1 з пробілом

Symbol	Frequency
и_	0.01944241
о_	0.01944063
_п	0.01673118
е_	0.01627276
_с	0.01424112
ст	0.01378449
я_	0.01352764
_н	0.01344202
_в	0.01315484
то	0.01252876

Таблиця частот біграм_1 без пробіла

Symbol	Frequency
ст	0.01650402
то	0.01518580
ен	0.01331289
но	0.01197148
ни	0.01184704
ко	0.01147794
на	0.01138724
по	0.01138514
ос	0.01114680
ов	0.01066381

Таблиця частот біграм_2 з пробілом

Symbol	Frequency
о_	0.01958511
и_	0.01924977
_п	0.01665983
е_	0.01635303
ст	0.01420545
_с	0.01401281
я_	0.01385227
_н	0.01349553
_в	0.01322084
а_	0.01270713

Таблиця частот біграм_2 без пробіла

Symbol	Frequency
ст	0.01642594
то	0.01493689
ен	0.01293322
ни	0.01194614
но	0.01192927
на	0.01153275
по	0.01146526
ко	0.01125013
ос	0.01122482
ов	0.01066801

Text with space:

Edited text written in text_edited_w_space.txt

Ентропія H1: 4.38810778332832 Надлишковість R1: 0.13010211345532208

Ентропія H2: 3.962854571765647 Надлишковість R2: 0.2144042519283479

Ентропія H2_2: 3.962447967693839 Надлишковість R2_2: 0.21448485706390774

Text without space:

Edited text written in text_edited_no_space.txt

Ентропія H1: 4.455010159894249 Надлишковість R1: 0.11683939547910716

Ентропія H2: 4.117433839898389 Надлишковість R2: 0.1837604789647077

Ентропія H2_2: 4.1153528455704755 Надлишковість R2_2: 0.18417301499553196

Text with space:

Edited text written in text_edited_w_space.txt

Ентропія H1: 4.38810778332832 Надлишковість R1: 0.13010211345532208

Ентропія H2: 3.962854571765647 Надлишковість R2: 0.2144042519283479

Ентропія H2_2: 3.962447967693839 Надлишковість R2_2: 0.21448485706390774

Text without space:

Edited text written in text_edited_no_space.txt

Ентропія H1: 4.455010159894249 Надлишковість R1: 0.11683939547910716

Ентропія H2: 4.117433839898389 Надлишковість R2: 0.1837604789647077

Ентропія H2_2: 4.1153528455704755 Надлишковість R2_2: 0.18417301499553196

Calculate redundancy

$0.513715532229392 < H^{(10)} < 0.3567859794231294$

$0.6692421767736474 < H^{(20)} < 0.5129398194470379$

$0.6002843539643062 < H^{(30)} < 0.4521141036497226$

CoolPinkProgram

Лабораторная работа №1

Произвольная часть текста:
ь_естественный_закон_и_как_только_кто_нибудь_начинает_говорить_мне_что_я_ег

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: е

Символ по счету: 1

Номер эксперимента: 50

Неравенство для энтропии:
 $2,40914871605956 < H < 3,18660852753104$

Двоичная таблица угаданных символов:

10000000000000000000000000000000	▲
00000000000000000000000000000000	■
10000000000000000000000000000000	
10000000000000000000000000000000	
10000000000000000000000000000000	▼

Поле ввода символов:
е

Продолжить Другой

Вероятности:

$q[1] = 0,44$
$q[2] = 0,08$
$q[3] = 0,04$
$q[4] = 0,04$
$q[5] = 0,02$
$q[6] = 0,06$
$q[7] = 0,04$
$q[8] = 0$
$q[9] = 0,02$
$q[10] = 0,02$
$q[11] = 0,04$
$q[12] = 0$
$q[13] = 0$
$q[14] = 0,04$
$q[15] = 0$
$q[16] = 0$
$q[17] = 0$
$q[18] = 0,02$
$q[19] = 0,02$
$q[20] = 0,02$
$q[21] = 0,02$
$q[22] = 0$
$q[23] = 0$
$q[24] = 0$
$q[25] = 0$
$q[26] = 0,02$
$q[27] = 0,04$
$q[28] = 0$
$q[29] = 0$
$q[30] = 0$
$q[31] = 0,02$
$q[32] = 0$

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

$H^{(10)}$

$$2.40914871605956 < H^{(10)} < 3.18660852753104$$

Лабораторная работа №1

Произвольная часть текста:
коны_добра_и_зла_законами_природы_они_подразумевали_под_этим_закон_человече

Использованные буквы:
_, т, н, в, с, ш, р, п, о, е, а,

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: 12

Номер эксперимента: 50

Неравенство для энтропии:
 $1,63863918745959 < H < 2,41299174943185$

Двоичная таблица угаданных символов:

10000000000000000000000000000000	▲
00001000000000000000000000000000	■
10000000000000000000000000000000	
00010000000000000000000000000000	
10000000000000000000000000000000	▼

Поле ввода символов:
к

Продолжить Другой

Вероятности:

$q[1] = 0,54$
$q[2] = 0,14$
$q[3] = 0,06$
$q[4] = 0,04$
$q[5] = 0,06$
$q[6] = 0$
$q[7] = 0,02$
$q[8] = 0,02$
$q[9] = 0$
$q[10] = 0$
$q[11] = 0$
$q[12] = 0,02$
$q[13] = 0,02$
$q[14] = 0$
$q[15] = 0,02$
$q[16] = 0$
$q[17] = 0$
$q[18] = 0$
$q[19] = 0$
$q[20] = 0$
$q[21] = 0$
$q[22] = 0$
$q[23] = 0$
$q[24] = 0$
$q[25] = 0$
$q[26] = 0,04$
$q[27] = 0$
$q[28] = 0$
$q[29] = 0,02$
$q[30] = 0$
$q[31] = 0$
$q[32] = 0$

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

$H^{(20)}$

$$1.63863918745959 < H^{(20)} < 2.41299174943185$$

Лабораторная работа №1

Произвольная часть текста:
мент_нас_не_интересует_насколько_обоснованны_все_эти_извинения_и_пояснения_

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: Ь

Символ по счету: 1

Номер эксперимента: 50

Неравенство для энтропии:
 $1.98026977879394 < H < 2.71433428621155$

Двоичная таблица угаданных символов:

10000000000000000000000000000000
10000000000000000000000000000000
00001000000000000000000000000000
00001000000000000000000000000000
10000000000000000000000000000000

Поле ввода символов:
Ь

Продолжить Другой

Вероятности:

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

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

$H^{(30)}$

$$1.98026977879394 < H^{(30)} < 2.71433428621155$$

Calculate redundancy

$$0.5137155322229392 < H^{(10)} < 0.3567859794231294$$

$$0.6692421767736474 < H^{(20)} < 0.5129398194470379$$

$$0.6002843539643062 < H^{(30)} < 0.4521141036497226$$

Труднощі, які виникли під час виконання роботи

Довго не міг зрозуміти, що саме мається на увазі під «біграми, що не перетинаються». Також не одразу зрозумів як користуватися CoolPinkProgram.

Висновок: в цій лабораторній роботі було засвоєно поняття ентропії на символ джерела та його надлишковості, вивченню та порівнянню різних моделей джерела відкритого тексту для наближеного визначення ентропії, набутто практичних навичок щодо оцінки ентропії на символ джерела, а також знайдено надлишковість джерела за допомогою CoolPinkProgram.