

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

Виконали:
студенти групи ФІ-93
Оржахівський Максим

Перевірив:
Чорний О.М.

ЗАГАЛЬНІ ВІДОМОСТІ

1. 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.

2. Постановка задачі

Написати програму для експериментальної оцінки ентропії на символ джерела відкритого тексту, порівняти різні моделі джерела відкритого тексту для наближеного визначення ентропії. Оцінити надлишковість російської мови у різних моделях відкритого тексту.

3. Хід роботи

1. Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.
2. 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.
3. За допомогою програми CoolPinkProgram оцінити значення $H(10)$, $H(20)$, $H(30)$.
4. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

ПРАКТИЧНА ЧАСТИНА

1. Частота з пробілом:

_: 0.132178
о: 0.088172
е: 0.0758254
и: 0.0751677
а: 0.0689439
н: 0.0590259
т: 0.0521213
р: 0.0481852
с: 0.0474497
в: 0.0384759
л: 0.036445
к: 0.0298749
м: 0.0281107
д: 0.0238019
п: 0.022629
у: 0.0182092
г: 0.0178796
я: 0.0174389
ы: 0.0158933
з: 0.0136238
й: 0.0127094
б: 0.0124841
ь: 0.011674
ч: 0.0109716
х: 0.00913112
ж: 0.00622381
ц: 0.00614595
ю: 0.00544521
ш: 0.0049035
ф: 0.00469643
э: 0.00331815
щ: 0.00284602

Частота без пробілу:

о: 0.101601
е: 0.0873743
и: 0.0866165
а: 0.0794447
н: 0.0680161
т: 0.0600598
р: 0.0555243
с: 0.0546767
в: 0.0443362
л: 0.0419959
к: 0.0344252
м: 0.0323922
д: 0.0274271
п: 0.0260756
у: 0.0209827
г: 0.0206028
я: 0.020095
ы: 0.018314
з: 0.0156988
й: 0.0146451
б: 0.0143855
ь: 0.013452
ч: 0.0126427
х: 0.0105219
ж: 0.00717175
ц: 0.00708203
ю: 0.00627457
ш: 0.00565035
ф: 0.00541174
э: 0.00382353
щ: 0.0032795

1)З перетином та пробілом:

[illegible]

	а	б	в	г	д	е	ж	з	и	к	л	м	н	о
а	2.31922e-05	0.00131533	0.00324029	0.000987327	0.0016102	0.00100721	0.000748778	0.0025577	0.000294873	0.000546674	0.00341257	0.00612938	0.00262404	0.00827963
б	0.000619564	2.98186e-05	1.93791e-05	0.00045117	0.0013527e-05	0.00045117	3.31318e-06	1.35272e-05	0.000907811	0.003445e-05	0.000911124	4.96077e-05	0.000381015	0.00429825
в	0.000331	0.00203e-05	0.20305e-05	0.61426e-05	0.00010270	0.00061159	0.000135	0.000272	0.000272	0.000272	0.00031424	0.00013330	0.00015100	0.00053550
г	0.00841213	0.0031318e-06	0.95767e-05	0.000261741	0.00188051	0.000188051	0.00028457	0.00028457	0.00028457	0.0031318e-06	0.00091124	4.96077e-05	0.00014267	0.00060838
д	0.00049007	0.93935e-06	0.000518566	2.31922e-05	0.00015961	0.000411828	0.000166222	2.65954e-05	0.000215688	0.0000185538	0.00094437	0.000125901	0.00155057	0.00296861
е	0.000282933	0.00050029	0.00191912	0.00148989	0.000382089	0.00115299	0.000616251	0.00111985	0.000175598	0.00333968	0.00216351	0.00059988	0.00449267	0.01028551
ж	0.000477998	3.6445e-05	3.31318e-06	0.62636e-06	0.000735526	0.000236561	0.93935e-06	3.31318e-06	0.00117949	0.00017949	5.6324e-05	3.31318e-06	0.000771971	4.30713e-05
з	0.00076061	0.00019937	0.00019937	0.00077877	0.00077877	0.937581e-05	0.00077877	0.00077877	0.00077877	0.00077877	0.000172789	0.00018081	0.00018081	0.00018081
и	0.000624081	0.000470741	0.00072407	0.00072407	0.00072407	0.00072407	0.00072407	0.00072407	0.00072407	0.00224631	0.000224631	0.000224631	0.000224631	0.000224631
к	3.31318e-06	2.98186e-05	0.298186e-05	2.98186e-05	1.93791e-05	0.0013527e-05	0.93935e-06	1.35272e-05	0.93935e-06	0.003445e-05	0.00612938	0.97581e-05	0.00295405	0.00192827
л	0.000544686	0.0000241862	0.62636e-06	3.31318e-06	0.000012938	0.000430713	0.62636e-05	0.000489688	0.93935e-05	0.00133521	6.29504e-05	0.000255115	0.0001206	0.00044839
м	0.00374389	1.35272e-05	3.31318e-06	0.61426e-05	0.97581e-05	0.00622215	0.000228609	0.00062304	0.00031439	0.000381015	1.35272e-05	0.000390955	0.000464839	0.000464839
н	0.00072733	0.000318065	5.30109e-05	0.93935e-06	0.62756493	3.31318e-06	0.00058486	0.00058486	0.000129214	0.00021983	0.000331318	0.000593059	0.000269198	0.000269198
о	0.000102709	0.000102709	0.000102709	0.000102709	0.000102709	0.000102709	0.000102709	0.000102709	0.000102709	0.000102709	0.000102709	0.000102709	0.000102709	0.000102709
а	0.00156051	2.31922e-05	0.00047118	0.00047118	0.00051702	0.00203098	0.00145117	0.001534	0.0000461135	0.0047374	0.0020111	0.00060333	0.0005341	0.00049117
б	0.00100456	0.000122588	0.000599685	0.00133852	0.000241983	0.00077794	0.000447279	0.93935e-06	0.00612121	0.000314752	0.000265504	0.00146442	0.00107016	0.00084446
в	0.00152075	0.00013527	0.00129877	2.31922e-05	0.000245175	0.00209111	0.0007794	0.000322524	0.000594384	0.00267682	0.00057318	0.00107021	0.000390624	0.000390624
г	0.00131922	0.00013527	0.00013527	0.00013527	0.00013527	0.00								

[illegible]


```
H1: 4.42399
H2: 8.07756
H2_peretyinless: 8.07759
H1_spaceless = 4.44862
H2_spaceless = 8.29619
H2_spaceless and peretyinless = 8.27957
```

Лабораторная работа №1

Произвольная часть текста:
али_сделать_для_такого_то_старого_своего_приятеля_обещали_и_не_сделали_что_

Использованные буквы:
л, н, й, __, е, в, я.

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: **t**

Символ по счету: **8**

Номер эксперимента: **50**

Неравенство для энтропии:
1,96763137009845 < H < 2,7079764393853

Двоичная таблица угаданных символов:

1000000000000000000000000000000000	▲
1000000000000000000000000000000000	
1000000000000000000000000000000000	
0100000000000000000000000000000000	
1000000000000000000000000000000000	▼

Вероятности:

```
q[1] = 0,48
q[2] = 0,16
q[3] = 0,06
q[4] = 0,02
q[5] = 0
q[6] = 0
q[7] = 0,08
q[8] = 0,02
q[9] = 0,02
q[10] = 0
q[11] = 0
q[12] = 0
q[13] = 0
q[14] = 0,02
q[15] = 0
q[16] = 0,02
q[17] = 0,02
q[18] = 0,02
q[19] = 0
q[20] = 0,02
q[21] = 0
q[22] = 0
q[23] = 0,02
q[24] = 0,02
q[25] = 0
q[26] = 0,02
q[27] = 0
q[28] = 0
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

Произвольная часть текста:
ть_естественный_закон_и_как_только_кто_нибудь_начинает_говорить_мне_что_я_е

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: 50

Неравенство для энтропии:
 $1,78681768823959 < H < 2,59092368847664$

Двоичная таблица угаданных символов:

0000000001000000000000000000000000
1000000000000000000000000000000000
0100000000000000000000000000000000
1000000000000000000000000000000000
0000000000010000000000000000000000

Вероятности:

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

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

Произвольная часть текста:
твенным_то_есть_законом_природы_сегодня_когда_мы_говорим_о_законах_природы_

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: д

Символ по счету: 1

Номер эксперимента: 50

Неравенство для энтропии:
 $1,28869407137585 < H < 2,00928370768122$

Двоичная таблица угаданных символов:

00000100000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
01000000000000000000000000000000
10000000000000000000000000000000

Вероятности:

$q[1] = 0,68$
$q[2] = 0,06$
$q[3] = 0,04$
$q[4] = 0,02$
$q[5] = 0,02$
$q[6] = 0,04$
$q[7] = 0$
$q[8] = 0$
$q[9] = 0,02$
$q[10] = 0,02$
$q[11] = 0$
$q[12] = 0,02$
$q[13] = 0$
$q[14] = 0$
$q[15] = 0$
$q[16] = 0$
$q[17] = 0,02$
$q[18] = 0$
$q[19] = 0$
$q[20] = 0$
$q[21] = 0$
$q[22] = 0$
$q[23] = 0$
$q[24] = 0,02$
$q[25] = 0$
$q[26] = 0$
$q[27] = 0$
$q[28] = 0$
$q[29] = 0,02$
$q[30] = 0,02$
$q[31] = 0$
$q[32] = 0$

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

4. Надлишковість російської мови:

$$R = 1 - \frac{H_{\infty}}{H_0}$$

$$H_0 = 5$$

$$H_1 = 4.42399$$

$$R = 1 - 4.42399/5 = 11,52\%$$