

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

КРИПТОГРАФІЯ

Комп'ютерний практикум

Робота № 1

*«Експериментальна оцінка ентропії на символ джерела
відкритого тексту»*

Виконав:

студент гр. ФБ-02

Шубін Д.Ю

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 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 $(10) H$, $(20) H$, $(30) H$.*
- 3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.*

Хід роботи:

- 1) Частоти букв у тексті з пробілами

а	0,06953
б	0,01352
в	0,03372
г	0,0184
д	0,02625
е	0,06564
ё	0,00001
э	0,00232
ж	0,00813
з	0,01596
и	0,05901
ы	0,01364
й	0,00757
к	0,02668
л	0,04597
м	0,02668
н	0,05415
о	0,09251
п	0,02464
р	0,0472
с	0,04492
т	0,04783
у	0,02546
ф	0,00228
х	0,00696
ц	0,00236
ч	0,01241
ш	0,00718
щ	0,00248
ъ	0,00024
ь	0,01626
ю	0,00389
я	0,01658
	0,15963

2) Частоти букв у тексті без пробілів

	а	0,08274
	б	0,01608
	в	0,04013
	г	0,02189
	д	0,03123
	е	0,07811
	ё	0,00002
	э	0,00276
0	ж	0,00967
1	з	0,019
2	и	0,07022
3	ы	0,01623
4	й	0,009
5	к	0,03175
5	л	0,0547
7	м	0,03174
3	н	0,06443
9	о	0,11009
0	п	0,02932
1	р	0,05617
2	с	0,05345
3	т	0,05691
4	у	0,0303
5	ф	0,00271
5	х	0,00828
7	ц	0,00281
3	ч	0,01477
9	ш	0,00855
0	щ	0,00295
1	ъ	0,00029
2	ь	0,01934
3	ю	0,00463
4	я	0,01973

3) Частота біграм у тексті з пробілами

[illegible]

4) Частота біграм в тексті з пробілами, з 2-им кроком

[illegible]

5) Частота біграм в тексті без пробілів

[illegible]

6) Частота біграм в тексті без пробілів, з 2-им кроком

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	
1	#																																		
2	#	0.00006	0.00159	0.00091	0.00234	0.00336	0.00191	0.00001	0.00001	0.00001	0.00012	0.00496	0.00115	0.0	0.00071	0.0058	0.01084	0.00463	0.00555	0.01682	0.00354	0.00684	0.00044	0.00585	0.00079	0.00026	0.00117	0.00017	0.00154	0.00039	0.00003	0	0	0.00096	0.00238
3	#	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	
4	#	0.00059	0.00029	0.00049	0.00009	0.00009	0.00058	0.0	0.0	0.00001	0.00012	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	
5	#	0.00021	0.00002	0.00002	0.00001	0.00016	0.00105	0.0	0.00001	0.0	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	
6	#	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	
7	#	0.00023	0.00031	0.00033	0.00073	0.00409	0.0162	0.0	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	
8	#	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	#	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
10	#	0.0145	0.00004	0.0001	0.00006	0.00019	0.00019	0.0	0.00001	0.00001	0.00001	0.00003	0.0016	0.0	0.00019	0.00003	0.00004	0.00009	0.00017	0.00015	0.00002	0.00008	0.00026	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	
11	#	0.00745	0.00022	0.00007	0.0006	0.00106	0.00062	0.0	0.00002	0.00007	0.00022	0.00033	0.00039	0.0	0.00036	0.00115	0.00043	0.00214	0.00019	0.00023	0.00049	0.00016	0.00008	0.00066	0.0	0.00003	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	
12	#	0.00129	0.00001	0.00001	0.00001																														
13	#	0.00007	0.00043	0.00143	0.00001	0.00017	0.00012	0.0	0.00001	0.00014	0.00029	0.00036	0.0	0.00147	0.00041	0.0182	0.01219	0.00036	0.00038	0.00042	0.0101	0.00091	0.00021	0.00009	0.00099	0.00091	0.00032	0.00008	0.00001	0.0	0.0	0.00008	0.00008		
14	#	0.00001	0.00002	0.00008	0.0																														

- 7) За допомогою скрипта Lab_1.py дізнаємось ентропію та надлишковість:

```
===Обрахунки для тексту з пробілами===  
H1=4.376010520712462  
Надлишковість = 0.13984422938075436  
H2=3.993079216410058  
Надлишковість = 0.21511383158747865  
H2(перехресна) = 3.993817939439508  
Надлишковість = 0.21496862698304198  
===Обрахунки для тексту без пробілів===  
H1=4.453531058572074  
Надлишковість = 0.11713261232283045  
H2=4.157189231404259  
Надлишковість = 0.17587937559229194  
H2(перехресна) = 4.156678632844339  
Надлишковість = 0.17598059658094556  
  
Process finished with exit code 0
```

- 8) Заходимо у програму *CoolPinkProgram*
9) Завдяки програмі оцінили значення H10 (10 символів)

11) Завдяки програмі оцінили значення H30 (30 символів)

Лабораторная работа №1

Произвольная часть текста:
ого_своего_приятеля_обещали_и

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: 58

Неравенство для энтропии:
 $2,14981146712401 < H < 2,94485835402027$

Двоичная таблица угаданных символов:

01000000000000000000000000000000	^
10000000000000000000000000000000	
00000000000000000000000000000000	
00100000000000000000000000000000	
10000000000000000000000000000000	^

Вероятности:

q[1]	= 0,4561403
q[2]	= 0,1228070
q[3]	= 0,0701754
q[4]	= 0
q[5]	= 0,0350877
q[6]	= 0,0350877
q[7]	= 0
q[8]	= 0,0350877
q[9]	= 0,0526315
q[10]	= 0,017543
q[11]	= 0
q[12]	= 0,035087
q[13]	= 0
q[14]	= 0,017543
q[15]	= 0
q[16]	= 0,017543
q[17]	= 0
q[18]	= 0
q[19]	= 0,035087
q[20]	= 0,017543
q[21]	= 0,017543
q[22]	= 0
q[23]	= 0
q[24]	= 0,017543
q[25]	= 0
q[26]	= 0
q[27]	= 0,017543
q[28]	= 0
q[29]	= 0
q[30]	= 0
q[31]	= 0
q[32]	= 0

Строка состояния:

Продолжить Другой

12) За знайденими значеннями знайдемо надлишковість, підстави значення у формулу, з скрипта Lab_1.py:

```
print('Розрахуємо надлишковість для виводу програми CoolPinkProgram у випадку значення H10 (10 символів)')
h10 = ((2.25353504037451+2.93707466332325)/2)
rh10 = rozr_nadl(h10, len(alf_a))
print_(rh10)

print('Розрахуємо надлишковість для виводу програми CoolPinkProgram у випадку значення H20 (20 символів)')
h20 = ((1.48883048810586+2.23221634558576)/2)
rh20 = rozr_nadl(h20, len(alf_a))
print_(rh20)

print('Розрахуємо надлишковість для виводу програми CoolPinkProgram у випадку значення H30 (30 символів)')
h30 = ((2.14981146712401+2.94485835402027)/2)
rh30 = rozr_nadl(h30, len(alf_a))
print_(rh30)
```

```
Lab_1_2 x
надлишковість = 0,1170007000074300
Розрахуємо надлишковість для виводу програми CoolPinkProgram у випадку значення H10 (10 символів)
0,4855071212835862
Розрахуємо надлишковість для виводу програми CoolPinkProgram у випадку значення H20 (20 символів)
0,6311700924188628
Розрахуємо надлишковість для виводу програми CoolPinkProgram у випадку значення H30 (30 символів)
0,495016675878587
```

Висновок: у ході лабораторної роботи, я на практиці розібрався з поняттям ентропії, надлишковості, реалізувавши їх на мові програмування Python. Ознайомився з вивченням

та порівнянням різних моделей джерела відкритого тексту для наближеного визначення ентропії, набув практичних навичок щодо оцінки ентропії на символу джерела.