МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ» ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

КРИПТОГРАФІЯ

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

«Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Виконали:

Студенти групи 3 курсу груп ФБ-96 та ФБ-94 Ігнатенко Артем Васюченко Георгій

Мета:

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.

Постановка задачі:

- 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 1 Н та 2 Н за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення 1 Н та 2 Н на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де ймовірності замінити відповідними частотами. Також одержати значення 1 Н та 2 Н 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:

Використані формули:

Ентропія

$$H(Z) = -\sum_{i=1}^{n} p_i \log p_i.$$

```
def H(freq, n):
    temp = []
    for f in freq.values():
        temp.append(f * math.log(f, 2))
    temp = sorted(temp)
    H = -sum(temp) / n
    return H
```

Надлишковість

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

$$H_0 = \log_2 m,$$

```
def R(value):
    R_ = 1 - value / math.log(33, 2)
    print(R_)
    return (R_)
```

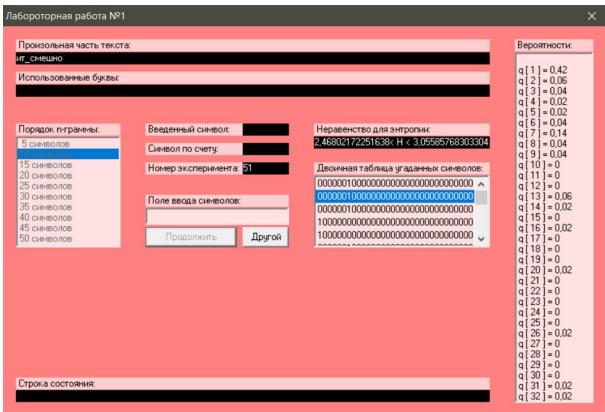
Задача 1:

Часто	та моног	рам без проб	ілів	τ	Іасто	та монс	ограм з пробілами
1	1 frequency_without_spaces				1	frequency_with_spaces	
2	0	0,11459	-	-	2		0,173396
3	e	0,08699			3	0	0,094724
4	а	0,07957			4	e	0,071907
5	н	0,06501			5	а	0,065772
6	и	0,06477			6	н	0,053739
7	т	0,06468			7	И	0,053542
8	С	0,05287			8	т	0,053463
9	В	0,04621		_	9	c	0,043703
10	л	0,04591		_	10	В	0,038196
11	p	0,04178		_	11	Л	0,037949
12	<u>г</u>	0,03299		_	12	р	0,034539
13		0,03198		_	13	к	0,027269
14	M	0,0314			14	A	0,026436
15	у	0,02962		_	15	M	0,025959
16	<u>у</u> П	0,02741		_	16	У	0,024483
17		0,02741		_	17	п	0,022658
18	Я	0,02230		_	18	ь	0,018978
19	<u>я</u> ч	0,02134		_	19	Я	0,01764
20	6	0,01808		_	20	4	0,014947
21		0,01737		_	21	6	0,014357
22	r	0,01649		_	22	<u>r</u>	0,013946 0,013635
	ы			_	24	ы	0,013633
23	3	0,01538		-	25	3 w	0,012713
24 25	ж й	0,0114		-	26	<u>ж</u> й	0,00942
				_	27	x	0,007025
26	Х	0,0085		_	28		0,006796
27	ш	0,00822		_	29	ю	0,004638
28	ю	0,00561		_	30	э	0,002912
29	9	0,00352			31	щ.	0,002469
30	щ	0,00299			32		0,002289
31	<u> </u>	0,00277			33	ф.	0,001027
32	ф	0,00124			34	ъ	0,0002
33	ъ	0,00024					,

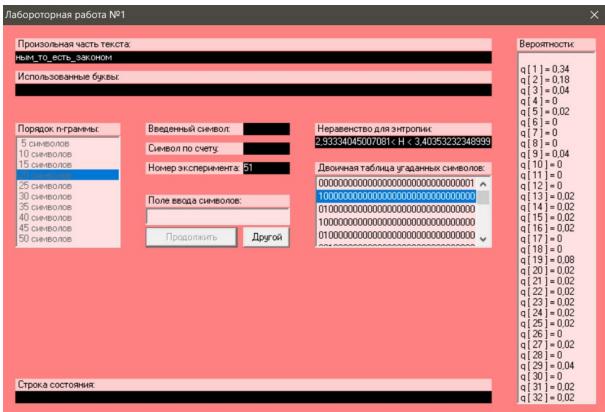
Частота біграм без пробілів										
A B C D F F G H I J 1 a 6 2 F F G H I J 2 a 0,00013 0,0001261 0,005171 0,000166 0,000001 1,886-06 0,002151 3 6 0,000741 2,786-06 8,8416-05 0 2,8416-05 0,0025651 0 2,476-06	6 DOMESS CONDISC CONDUCTS COND									
4 B 0,006879 0,000216 0,000479 0,000272 0,001355 0,005435 0 4,886-05 5 r 0,001207 2,888-05 0,00145 2,218-05 0,001383 0,000237 0 9,488-06 6 A 0,000597 6,688-05 0,001175 2,256-05 9,955-05 0,005306 2,376-06 2,888-05 7 e 0,000284 0,000597 0,001375 0,000482 0,0004524 0,0005212 0 0,000185	0 0.00045 3.551-06 0.000457 0.000255 0.000255 0.000255 0.000257 0.000353 0.001214 0.000351 0.000134 0.000355 0.000355 0.000355 0.000355 0.000355 0.000355 0.000355 0.000355 0.000355 0.000355 0.000355 0.000355 0.000355 0.000355 0.000355 0.000355 0.000355 0.000355 0.00035	0,000264 1,18E-05 0,000524								
8 8 0 1.18E-06 5.78E-06 1.18E-06 7.11E-06 3.55E-06 0 9 m 0,001465 6.5E-05 5.78E-06 1.07E-05 0.00097 0.003293 0 1.54E-05 10 s 0 0 0 0 0 0 0 0 11 u 0,000334 0.001668 0.005129 0.000998 0.003126 0.003447 0 0.000442		0 2,96E-05 0 0,001632								
12 A 9,726-50 (0,002399 (0,000574 (0,000241 (0,00073 9,722-65 0 0,000108 13	0 0.000055 0 0.000055 0.0000347 0.000086 0.000098 0.0000037 0.000055 0.000025 0.000252 0.000252 0.000252 0.000055 0.000025 0.000055 0.0000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.0000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.0000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.0000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.0000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.0000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.0000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.0000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.0000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.0000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.0000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.000005 0.0000005 0.0	0,000566								
15 H 0,01367 0,000275 0,00064 0,000109 0,000531 0,012221 2,37F-06 2,61E-05 17 0 0,000257 0,005644 0,012579 0,005399 0,007322 0,002844 0 0,000075 18 n 0,00055 4,78E-06 2,37F-06 1,18E-06 2,37F-06 0,00034 0 0 19 p 0,000579 0,000567 0,000504 0,000079 0,000393 0,005107 0 0,000284 00 c 0,000265 0,000265 0,000379 3,94E-05 0,00047 0,000539 7,11E-05 0,00054	0 CORPUSE LIMITING CONCENTRY SAME OF CONCENTRY	0,001325 0,00069 0,001171								
0	0 000599 0 000041 (00039) 000390 (000390 000	0,000579								
25	0 0,000174									
29	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0,000834								
8.5 N	0 0,000395 1,18F-64 (0,000361 6,28F-65 0,000394 0,000312 0,000313 0,000315 0,000317 0,000697 8,12F-65 1,5F-65 1,5F-65 1,2F-65 1,0F-65 0,000393 9,12F-65 0,000395 0,00									
Частота біграм з пробілами										
2 0,0062195 0,0025154 0,0068085 0,0166652 0,0009565 0,0060958 0,0060958 0 0,0022383 0 0,0022383 0 0,0062585 0,0060958 0,0061958 0,0061958 0,0061958 0,0061958 0,006185 0,00618	5 0 0.000117 0.00019% 0.00047% 0.00019% 0.00047% 0.000197 0.0011199 1.2714-05 0.0007481 0.000149% 0.00041% 0.00049% 0.000199 0.00	0,0006034								
6 r 0,0001025 0,0000984 0 5,8778-06 3,7795-07 0,0010951 0,0000729 0 4,877-06 7 7 A 0,00010571 0,00010571 0,00010571 0,00010571 0,00010571 0,00010571 0,00010571 0,00010571 0,0001057 0,000	0 0,000234 0 0,0001369 0,000641 7,001560 0,0015925 0,0015925 8,13150 0,0013959 0,0000364 0,0002731 0,0015952 0 4,500560 0,00002315 0,5007565 6,661560 1,5905-06 0 0,00002311 0,001595 0,001597 0,0005180 0,000137 0,0005180 0,000159 0,001590	0,0004241 0,0002135 0								
12 # 0 0 0 0 0 0 0 0 0 0 0 0 0 10 0 12 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0								
15 A 0,0001996 0,0003967 2,54710 5 1,7710 5 0,0001111 2,00710 5 0,0001096 0 0,0001096	6 0 0,0002941 0 0,0002915 4,00046 5 7,7004 0,000295 0,000297 3,001415 0 0,000295 0,000295 0,000295 0,000295 0	0,0003369 0,0017739 0,0006083								
35 m 3.536-65 0.0007856 0 0 0 0 0 0.0003892 0 0 0 0 0 0.0003892 0 0 0 0 0 0.0003892 0 0 0 0 0 0 0.0003892 0 0 0 0 0 0 0.0003892 0 0.00	0 0,00037400 8,7906-07 0,0003399 4,5006-07 0,0003311 0,0005041 0,0007054 8,7906-07 0,0003310 0,0003130 0,0	0,0035399								
36	0 0 0,000,000 0 1,3554 0,7364 0 1,3144 0 0,000,000 1,754 7 7,446 0 1,595 4 0,177 4 0,000,000 1,077 4 0 0 1,575 4 7,755 4 7 0 0 2,355 4 0,000,000 0 0 0 0,000,000 1 0 0 0 0,755 4 7,755 4 7 0 0 0 0 0,000,000 0 0 0 0,000,000 0 0 0 0,000,000 0 0 0 0,000,000 0 0 0 0,000,000 0 0 0 0,000,000 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0 0,000,000 0 0 0 0 0 0,000,000 0 0 0 0 0,000,000 0 0 0 0 0 0,000,000 0 0 0 0 0 0,000,000 0 0 0 0 0 0,000,000 0 0 0 0 0 0,000,000 0 0 0 0 0 0,000,000 0 0 0 0 0 0,000 0 0	0 0								
28		0 0,0001675 2,938E-06								
22	0 0 1,273E-05 4,114E-05 4,697E-06 1,959E-06 2,351E-05 0 4,897E-06 9,795E-07 6,856E-06 0,002724 0 8,815E-05 0 0 0 0 0 0 0 0 1,273E-05 0	0								
Частота біграм без пробілів, крок 2										
A B C D E F G H 1 1	C M N O P O R S T V W X Y A AR AC AC AR AC AC AR AC AC	AH 1 0,002718 5 0,000739								
4 B 0,006915 0,006225 0,000502 0,000275 0,001441 0,006993 0 5,21E-05 5 r 0,001159 4,005-05 0,001054 2,13E-05 0,000150 0,000205 0 0,9,48E-06 6 A 0,006242 7,82E-05 0,001152 2,13E-05 0,000104 0,005055 0 3,08E-05 7 e 0,000210 0,0002454 0,000345 0,000	0 000411 4344-06 0000614 0000704 0000705 0000151 0007714 0001118 0000705 000000705 0000705 0000705 0000705 0000705 00000705 00000705 00000705 0000705 0000705 0000705 0000705 0000705 0000705 0000705 0000705	6 0,00027 6 1,18E-05 5 0,00051								
8	0 0 0 4742-66 0 2.576-06 54845-66 4742-67 71115-66 71115-									
12	0 0,077056 0 0,0007705 0,000152 0,000161 0,000152 0,000705 0,000161 0,0007059 0,00016 0,0002256 0,000047 0,000559 0,00056 0,000182 0,0007059 0,00056 0,000182 0,0007059 0,00056 0,000182 0,0007059 0,00056 0,000182 0,000569 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,00018 0,0007059 0,00056 0,0007059 0,00056 0,0007059 0,00056 0,0007059 0,00056 0,0007059 0,0007059 0,00056 0,00056 0,0007059 0,00056 0,0007059 0,00056 0,0007059 0,00056 0,0007059 0,00056 0,0	0 0,000128 8 0,001595 5 0,000547								
15 H 0,01194 0,00222 0,00266 0,00202 0,00295 0,012245 0 2,378 05 17 0 0,000254 0,005633 0,01245 0,005444 0,00715 0,00283 0 0,00806 18 n 0,00972 7,141-06 0 2,378-06 2,378-06 0,00342 0 0 19 p 0,008630 0,000166 0,000481 0,0022 0,000417 0,006005 0 0,000225 20 c 0,002105 0,002160 0,003589 2,756 0,0002479 0,005005 0 0,000251	0 0000944 2377-00 0000014 8377-00 0000014 8377-00 0000014 0000137 0000017 0000015 0000015 0000015 000015 000015 000015 000015 000015 000015 000015 000015 0000015 0000015 0000015 0000015 0000015 0000015 0000015 0000015 00000010 0000015 0000015 0000015 00000000	0,001313 0,000659 6 0,001199								
00 00 00 00 00 00 00 0	0 0,004299 0 0,000737 0,000313 0,000251 0,001901 0,018218 0,000628 0,003808 0,001787 0,000486 0,00218 2,13E-05 5,92E-05 8,77E-05 0,000685 3,08E-05 3,79E-05 0,001676 0,007816 0,000128 0,000118									
15 4 0,00064 1,881-05 8,298-05 2,377-06 1,98-05 0,000889 0 0 0 0 0 0 0 0 0	0 0,00018 0 4,27E-05 9,48E-06 9,48E-06 3,00E-05 0,000224 8,06E-05 1,42E-05 2,17E-05 4,08E-05 0,000209 0 0 0 2,61E-05 0 0 0,000219 0 4,74E-06 0 0,00127 0 0,000611 1,9E-05 9,48E-06 0,001102 0,000218 6,4E-05 0,000137 4,74E-05 0,005098 0,000827 0 4,74E-06 0 1,42E-05 0,000175 0 0 0 0,000306 7,11E-06 0	0 7,11E-06 0 1,42E-05 0 2,37E-06								
29 * 0 0 0 0 0 0 0 4,272-05 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 0 0 0 0	0 0,000213 0 6,87E-05 3 0,000799 0 0								
33 e 8,29E-05 0,000434 0,000322 7,35E-05 0,000495 5,92E-05 0 9,34E-05 34 a 0,000315 0,000438 0,001664 0,000277 0,001294 0,000441 0 0,000306	0 0,000255 0 0,000277 5,00F45 0,000215 0,000306 0,00022 0,000322 0,00035 0,000611 7,555-65 2,138-65 1,18-65 7,118-66 0,000399 7,268-65 0,000306 0 0 6,64-65 4,058-66 0,000316 9,000216 0,000277 0,000114 0,00078 0,000217 0,000218 0,000318 0,00018 0,									
Частота біграм з пробілами, крок 2										
A B C D E F G H 1	1	AD 0								
4 6 0,000437 0,000613 1,96E-06 4,7E-05 0 1,76E-05 0,002114 5 8 0,00585 0,005801 1,96E-06 2,94E-05 1,57E-05 0,000866 0,004937 6 r 0,001056 0,001015 0 1,96E-06 1,96E-06 0,001113 0,000174	0 1,96E-06 0 0,00078 0 0,000145 0,000697 4,7E-05 0,000247 0,0019 0 0,001171 0,000188 3,92E-06 0,001218 0 5,29E-05 0 1,57E-05 7,84E-06 0,00018	0,000155 3,92E-06 0								
8 e 0,018865 3,92E-06 0,001305 0,001128 0,003221 0,002672 0,001718 9 8 5,09E-05 0 0 0 0 0 0 0 10 ж 0,000678 0,001152 1,96E-05 0 3,92E-06 0,000556 0,00438	0 0,000768 0 8,216-05 0,002145 0,002121 0,005536 0,004229 0,004432 0,000243 0,001111 0,005738 0,00445 0,005789 0,000163 5,886-05 0,000537 0,00039 0,00116 0,001027 0,00096 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0								
11 a 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0,00237 0 0,000533 0,00978 0,00272 0,00688 0,00246 0,00268 0,00268 0,00053 0,00019 0,00018 0,00019 0,00487 7,84E-06 1,74E-50 0,00713 0,00618	0 0								
15 n 0,00696 0,00307 3,33E-05 1,18E-05 0,000131 1,96E-05 0,003628 16 m 0,007278 0,00276 5,88E-06 1,96E-06 8,8E-05 1,96E-06 0,003386 17 н 0,004139 0,00977 1,57E-05 1,37E-05 2,94E-05 0,002278 0,009895 18 o 0,023626 7,84E-06 0,003346 0,007834 0,00465 0,009862 0,001708	0 0,003393 0 0,005182 0 0,000247 3,315-05 5,886-06 0,000244 0,005184 7,584-05 0 0,0001379 0,000108 0,000154 1,584-06 0 0 0,000188 0 1,584-06 0 0 0,000188 0 0,0054-0,0054-0,000134 1,584-06 0,00014 0,000131 1,584-06 0,00014 0,000131 1,584-06 0,00014 0,000131 1,584-06 0,00014 0,000131 1,584-06 0,00014 0,000131 1,584-06 0,00014 0,000131 0,00014	0								
19 n 3,72E-05 0,000758 0 0 0 0 0,002905 20 p 0,000733 0,007217 0,000118 0,000382 0,000149 0,00027 0,005027	0 0 0 0,000772 0 6,46E-05 0,000625 0 9,4E-05 0,000812 2,74E-05 0,006498 2,74E-05 0,000178 0 0 7,84E-06 1,76E-05 3,92E-06 0 0 0,000237 0 0,005187 1,66E-06 0,000208 5,88E-05 0,000821 0,00697 8,62E-05 2,35E-05 0,000119 0,000643 0,002549 0,000302 5,29E-05 1,57E-05 7,44E-05 0,000214 7,84E-06	0 1,57E-05								
23 Y 0,005834 0,005/03 1,962-06 0,002039 1,372-05 7,448-05 0,005844 23 Y 0,006759 3,722-05 0,00068 0,000664 0,00143 0,00208 0,000165 24 ф 1,372-05 0,00021 0 0 0 0 4,312-05	0 0 0,00183 0 7,846-06 0,00023 0,000639 0,001444 0,001593 0,000523 0 0,000662 0,000662 0,000683 0,00169 0,00169 0,000663 0,00169 0,000662 0,000662 0,000683 0,00169 0,00169 0,000683 0,00169 0,00088 0,00169 0	0,792-06								

Діаграми також ϵ в .xlsx файлах з кращою якістю зображення.

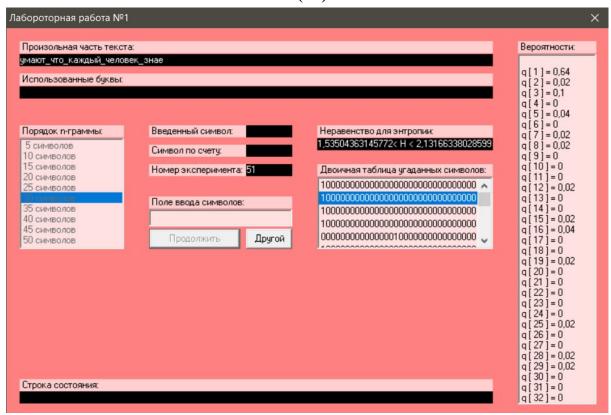
H(10)



H(20)



H(30)



Задача 3:

2.468021 < H(10) < 3.055857	0.510739 > R(10) > 0.394207			
2.933340 < H(20) < 3.403532	0.418495 > R(20) > 0.325284			
1.535043 < H(30) < 2.131663	0.695693 > R(30) > 0.577419			

Висновки:

Здобуто навички вимірювання частот повторювання різних символів в довільному тексті, оцінки ентропії, надлишковості на прикладі російської мови, підрахування частот монограм та біграм у довільному тексті.