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

# КРИПТОГРАФІЯ КОМП'ЮТЕРНИЙ ПРАКТИКУМ No1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконали: Бойко Т. Я.

Хандрос А. В.

Група: ФБ-02

# 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.

#### Завдання

Створити програму для розрахування частот букв та біграм у тексті достатньої довжини російською мовою, а також ентропії за її безпосереднім означенням. Застосувати цю програму обрахувавши частоти букв та біграм, ентропію на тексті. Оцінити надлишковість мови на основі отриманих результатів.

# Хід роботи

## 1. Частота букв

	freq with spaces	freq without spaces
	0,1628	
0	0,089	0,1062
а	0,0801	0,0957
е	0,0699	0,0835
н	0,0545	0,0652
И	0,0529	0,0632
т	0,0496	0,0592
л	0,046	0,055
С	0,0437	0,0522
р	0,0372	0,0444
В	0,0331	0,0395
К	0,0296	0,0353
у	0,028	0,0335
M	0,0264	0,0316
п	0,0234	0,0279
Д	0,0227	0,0271
Я	0,02	0,0239
ь	0,0186	0,0222
ы	0,0154	0,0184
3	0,0152	0,0182
г	0,0126	0,0151
б	0,0126	0,015
ч	0,0119	0,0142
й	0,0087	0,0104
ж	0,0082	0,0098
ш	0,0068	0,0082
х	0,0061	0,0073
ю	0,0044	0,0052
щ	0,0031	0,0038
ц	0,0029	0,0035
ф	0,0025	0,003
9	0,0019	0,0023
ъ	0,0002	0,0002

# 2. Частота біграм (повний список в .xlsx файлі)

	перехр. з пробілом		перехр. без пробіла
а	0,026124567	ла	0,017653808
0	0,019833486	то	0,01494615
е	0,018268913	на	0,013452062
П	0,016211607	но	0,011604409
Н	0,015639748	ПО	0,011541151
И	0,01522835	СТ	0,011407481
ла	0,014670046	ен	0,011000071
Я	0,013480932	ал	0,01056894
С	0,013423872	не	0,01042661
В	0,013007115	ко	0,009609907
ь	0,012385447	ОС	0,009388128
то	0,012120008	ра	0,009244292
O	0,011551302	ac	0,009167856
на	0,011167015	ро	0,0089457
ПО	0,009655088	ОВ	0,008879054
но	0,009569025	ОН	0,008667442
СТ	0,009322817	ОТ	0,008613597
не	0,00866521	ни	0,008476539
Т	0,008502858	ОЛ	0,007687323
И	0,008450211	ат	0,007500562
ал	0,008161129	ка	0,007495291
	з пробілом		без пробіла
а	0,013061338	ла	0,008785673
a O	0,013061338 0,009918004	ла то	0,008785673 0,007488889
	0,013061338 0,009918004 0,009088903		0,008785673 0,007488889 0,006760296
O	0,013061338 0,009918004 0,009088903 0,008033139	то	0,008785673 0,007488889 0,006760296 0,005757585
0 e	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703	то на	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005
о е п	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772	то на по	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994
о е п н	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772 0,007339909	то на по но	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994 0,005483092
о е п н	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772 0,007339909 0,006711306	то на по но ст	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994 0,005483092 0,005292565
о е п н и	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772 0,007339909 0,006711306 0,006673161	то на по но ст ен	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994 0,005483092 0,005292565 0,005190524
о е п н и ла	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772 0,007339909 0,006711306 0,006673161 0,006524994	то на по но ст ен ал	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994 0,005483092 0,005292565 0,005190524 0,004830181
О е П Н И ла Я	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772 0,007339909 0,006711306 0,006673161 0,006524994 0,00617507	то на по но ст ен ал не	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994 0,005483092 0,005292565 0,005190524 0,004708184
О е П Н И ла я С	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772 0,007339909 0,006711306 0,006673161 0,006524994 0,00617507 0,006050862	то на по но ст ен ал не ко ос ас	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994 0,005483092 0,005292565 0,005190524 0,004830181 0,004708184 0,004621958
О е П Н И Ла Я С В Ь	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772 0,007339909 0,006711306 0,006673161 0,006524994 0,00617507 0,006050862 0,005775336	то на по но ст ен ал не ко ос ас ра	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994 0,005483092 0,005292565 0,005190524 0,004830181 0,004708184 0,004621958 0,004620828
О е П Н И Ла Я С В Ь ТО О	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772 0,007339909 0,006711306 0,006673161 0,006524994 0,00617507 0,006050862 0,005775336	то на по но ст ен ал не ко ос ас ра	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994 0,005483092 0,005292565 0,005190524 0,004830181 0,004708184 0,004621958 0,004620828 0,004503726
О е П Н И Ла Я С В Ь ТО О На ПО	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772 0,007339909 0,006711306 0,006673161 0,006524994 0,00617507 0,006050862 0,005775336 0,005559706 0,004866162	то на по но ст ен ал не ко ос ас ра ро	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994 0,005483092 0,005292565 0,005190524 0,004830181 0,004708184 0,004621958 0,004620828 0,004503726 0,0044544
О е П Н И Ла Я С В Ь ТО О На ПО НО	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772 0,007339909 0,006711306 0,006673161 0,006524994 0,00617507 0,006050862 0,005775336 0,005559706 0,004866162 0,004754564	то на по но ст ен ал не ко ос ас ра ро ов	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994 0,005483092 0,005292565 0,005190524 0,004830181 0,004708184 0,004621958 0,004620828 0,004503726 0,0044544 0,004305669
О е П Н И Ла Я С В Ь ТО О На ПО НО СТ	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772 0,007339909 0,006711306 0,006673161 0,006524994 0,00617507 0,006050862 0,005775336 0,005559706 0,004866162 0,004754564 0,00467323	то на по но ст ен ал не ко ос ас ра ро ов он	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994 0,005483092 0,005292565 0,005190524 0,004830181 0,004708184 0,004621958 0,004620828 0,004503726 0,0044544 0,004305669 0,004295879
О е П Н И Ла Я С В Ь ТО О На ПО НО СТ Не	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772 0,007339909 0,006711306 0,006673161 0,006524994 0,00617507 0,006050862 0,005775336 0,005559706 0,004866162 0,004754564 0,00467323 0,004363342	то на по но ст ен ал не ко ос ас ра ро ов он	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994 0,005483092 0,005292565 0,005190524 0,004830181 0,004708184 0,004621958 0,004620828 0,004503726 0,0044544 0,004305669 0,004295879 0,00421643
О е П Н И Ла Я С В Ь ТО О На ПО НО СТ Не Т	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772 0,007339909 0,006711306 0,006673161 0,006524994 0,00617507 0,006050862 0,005775336 0,005559706 0,004866162 0,004754564 0,00467323 0,004246385	то на по но ст ен ал не ко ос ас ра ро ов он от ни	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994 0,005483092 0,005292565 0,005190524 0,004830181 0,004621958 0,004620828 0,004503726 0,0044544 0,004305669 0,004295879 0,00421643 0,003830483
О е П Н И Ла Я С В Ь ТО О На ПО НО СТ Не	0,013061338 0,009918004 0,009088903 0,008033139 0,007796703 0,007603772 0,007339909 0,006711306 0,006673161 0,006524994 0,00617507 0,006050862 0,005775336 0,005559706 0,004866162 0,004754564 0,00467323 0,004363342	то на по но ст ен ал не ко ос ас ра ро ов он	0,008785673 0,007488889 0,006760296 0,005757585 0,005738005 0,005673994 0,005483092 0,005292565 0,005190524 0,004830181 0,004708184 0,004621958 0,004620828 0,004503726 0,0044544 0,004305669 0,004295879 0,00421643

В коді на жаль  $\epsilon$  помилка, яку не змогли знайти( через що у випадку із перехресними біграмами отриму $\epsilon$ мо не дуже адекватні значення

3. Ентропія

H1 with spaces: 4.365662742377314

H1 without spaces 4.450110269499657

H2 cross bi with spaces: 7.9606401833399225

H2 cross bi without spaces: 8.29965726118714

H2 bi with spaces: 4.48066352093309

H2 bi without spaces: 4.650355183976171

4. Надлишковість

R: 0.14187820558029463 – аналізуючи частоту букв з пробілами

R: -0.5647564280554915 – аналізуючи частоту перехресних біграм з пробілами

R: 0.11927346483932577 – аналізуючи частоту біграм з пробілами

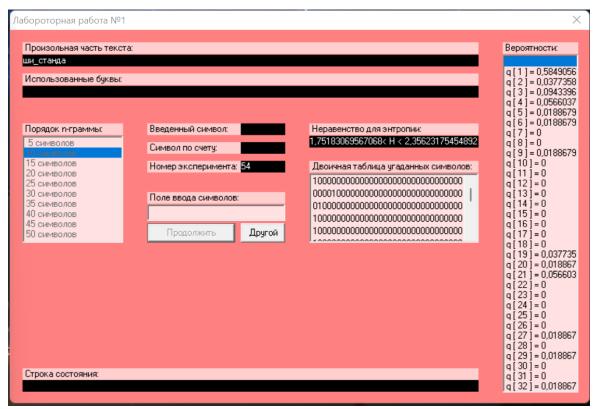
R: 0.11781074908048172 – аналізуючи частоту букв без пробілів

R: -0.6453229198202877 – аналізуючи частоту перехресних біграм без пробілів

R: 0.07811422463405693 – аналізуючи частоту біграм без пробілів

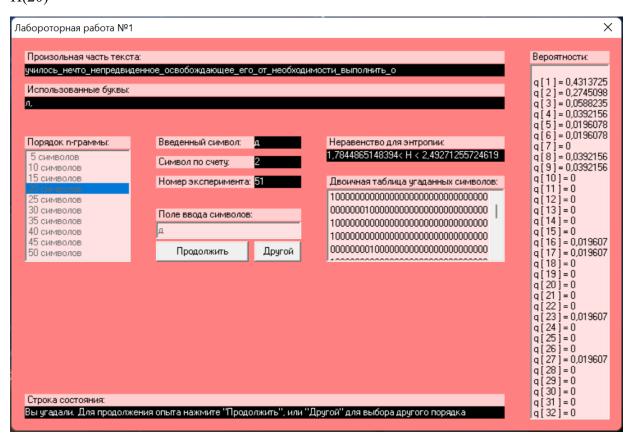
#### 5. CoolPinkProgram

#### H(10)



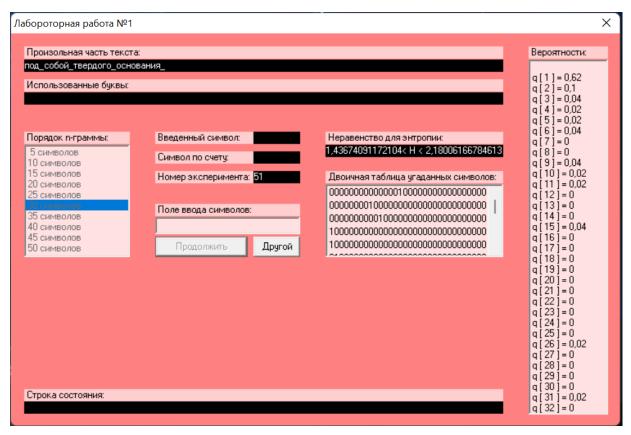
#### 1.752<H<2.356

## H(20)



### 1.784<H<2.492

# H(30)



1.437<H<2.18

#### 6. Висновок

Виконуючи лабораторну роботу ми дослідили ентропію та надлишковість російської мови на різних джерелах тексту. В теорії, результати не повинні сильно відрізнятися для різних джерел тексту, але на даному етапі код містить помилку, через яку перехресні біграми у тексті з пробілами обробляються некоректно (помилково обчислюється ентропія, а отже і надлишковість). Відловити помилку не зламавши всього іншого не вдалось.