

ФБ – 03 Серховець Анастасія, КІП №1

Мета роботи: Засвоєння понять ентропії та символ джерела та його надлишковість, вивчення та порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії та символ джерела

Виконавши код я отримала різні частотні таблиці, тут я навела декілька для прикладу, інші можна проглянути у вкладених файлах:

plain_spaces... onograms.csv	
1 ;0.15548635957624116	
2 а;0.06754592134612794	
3 б;0.015114996731487996	
4 в;0.04181200317370021	
5 г;0.01677669825396588	
6 д;0.024241879868461106	
7 е;0.0747466279435321	
8 ж;0.009186764272918257	
9 з;0.015045135406218655	
10 и;0.06068953128040839	
11 й;0.012071038987609596	
12 к;0.02761019376537573	
13 л;0.04022016297649166	
14 м;0.026093204988098626	
15 н;0.062356222897548365	
16 о;0.098404666736528	
17 п;0.02176679291606162	
18 р;0.037685194888147026	
19 с;0.047665384212338506	
20 т;0.04950173904798974	
21 у;0.02219594105700185	
22 ф;0.00141718688403519	
23 х;0.008203715624485396	
24 ц;0.003298452571645284	
25 ч;0.011721732361262893	
26 ш;0.006052984825122133	
27 щ;0.0028742945253671464	
28 ь;0.015514204304455655	
29 ю;0.00593821264789393	
30 я;0.01876275592947998	
31	

	;0.00267467739198994
а;0.0023453327877523728	
б;0.006636792782363098	
в;0.015319514166808053	
г;0.00457090390123654	
д;0.00610784538767852	
е;0.00413177762253116	
ж;0.0021856505553947644	
з;0.0036427509256579406	
и;0.010808491102705617	
к;0.008103873292148624	
л;0.00267467739198994	
м;0.005838381620575056	
н;0.013622890448008464	
о;0.011018074032674977	
п;0.015758640305791476	
р;0.004680685435982395	
с;0.015139871655405743	
т;0.00877254264014611	
у;0.003692651623269693	
ф;0.0005888282318186808	
х;0.001317378416950269	
ц;0.0004491062785057735	
ч;0.003572889949001487	
ш;0.00045908641802812404	
щ;4.990069761175261e-05	
ю;9.980139522350522e-06	
я;0.0012175770217267637	

plain_spaces_no_repeat_frequencies_bigrams.csv	
1 а;0.0024171689793276428	
2 б;0.0065435494963388486	
3 в;0.015231198673085481	
4 г;0.004480359237833246	
5 д;0.005865933087908087	
6 е;0.004126380517011206	
7 ж;0.0024070553015898702	
8 з;0.003600469274647033	
9 и;0.011013795056434322	
10 к;0.008343784133662366	
11 л;0.0027509203446741373	
12 м;0.0061086613536146285	
13 н;0.013795056434321776	
14 о;0.011034022411909867	
15 п;0.01565597313807193	
16 р;0.0045916096929487435	
17 с;0.015160402928921074	
18 т;0.008738217565435495	
19 у;0.003661151341073668	
20 ф;0.0006371616974796715	
21 х;0.001112504551154982	
22 ц;0.0005056838868886282	
23 ч;0.004035357417371253	

Також я порахувала ентропію біграм та монограм, ось результат:

```
Ентропія монограм для тексту з пробілами: 0.8570652589159388
Ентропія монограм для тексту без пробілів: 0.8685262553268784
Ентропія біграм для тексту з пробілами та повторами: 1.5673233037648688
Ентропія біграм для тексту без пробілів та з повторами: 1.620096162614525
Ентропія біграм для тексту з пробілами та без повторів: 1.5622450535397265
Ентропія біграм для тексту без пробілів та без повторів: 1.6125237431325077
```