

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

Криптографія
Комп'ютерний практикум №1

Виконав
студент гр. ФБ-03 Антіпов Данило
Перевірив
Чорний Олег Миколайович

Київ — 2022

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

Хід роботи

1. Підрахуємо частоту букв, а також H_1 та H_2

Частота літер з пробілами			Частота літер без пробілів	
0		0.1593991530412655	г	0.023834750548963973
1	г	0.020035511498509278	л	0.05394192277454976
2	л	0.045343625970869175	а	0.08386384789857496
3	а	0.0704960215727606	в	0.041685116073122905
4	в	0.03504054387664027	о	0.11020621165159086
5	о	0.09263943485444083	з	0.0182468395037185
6	з	0.015338308741145864	ы	0.017560069761347514
7	ы	0.014761009514043183	ш	0.008837642211037113
8	ш	0.00742892952771606	е	0.07880795749256753
9	е	0.0662460358153402	н	0.06430107624047784
10	н	0.05405153914810384	и	0.06914461021298898
11	и	0.058123017907670105	т	0.05749324525812602
12	т	0.04832887065838698	м	0.033620089822253146
13	м	0.028261075979414726	р	0.05381767166984448
14	р	0.045239180387018366	д	0.030508164427134634
15	д	0.02564518885660571	э	0.002671398751163442
16	э	0.0022455800527924953	п	0.027756567235661424
17	п	0.02333219392696405	я	0.01983047631096211
18	я	0.01666951518258987	с	0.05205782420501884
19	с	0.04375985111756775	ь	0.018632017928304854
20	ь	0.015662090051083384	б	0.0161074613917931

21	б	0.013539945688296397	к	0.03287232408302685
22	к	0.027632503465694375	у	0.02923402582615689
23	у	0.024574146869481	х	0.00818362957808843
24	х	0.0068791659545376855	й	0.00976161860784544
25	й	0.008205624869443021	щ	0.003507269819180756
26	щ	0.002948213980516151	ч	0.01389240306154722
27	ч	0.011677965779828709	ж	0.009698363499995482
28	ж	0.00815245257220988	ю	0.005544987936347288
29	ю	0.004661121555669496	ц	0.0025516658684474488
30	ц	0.00214493249017262	ф	0.0018287503501622042
31	ф	0.0015372490932224312		

H1 з пробілами: 4.37608270769908
 Надлишковість з пробілами: 0.12478345846018402
 H1 без пробілів: 4.453020149641895
 Надлишковість без пробілів: 0.10939597007162105

Таблиця значень:

	H1	
	з пробілами	без пробілів
H	4.3760827	4.4530201
R	0.1247834	0.1093959

Труднощі що виникли:

На жаль, не вдалося реалізувати обрахунок частоти біграм, ентропії та надлишковості для них. В основному, алгоритм обрахування ентропії та надлишковості мені повністю зрозумілий, приклади розрахунків розміщені нижче.

2. За допомогою CollPinkProgram оцінити значення $H^{(10)}$, $H^{(20)}$, $H^{(30)}$

Для обчислення надлишковості використаємо формулу з методичних вказівок:

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

$$H_0 = \log_2 m,$$

m - к-ть букв у алфавіті

Приклад для H1:

$$R = 1 - \frac{4.3760827}{\log_2(32)} = 0.1247834$$

[illegible]

$$1,3195163 < H(10) < 2,0549443$$

$$R = 0,6625539$$

[illegible]

$$1,5248948 < H(20) < 2,2311574$$

$$R = 0.62439478$$

