КРИПТОГРАФІЯ КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

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.

Хід роботи

Мова програмування для виконання практикуму -- Python 3(v 3.11), операційна система -- Fedora 38.

Було написано код, який рахує частоту з'явлення букв у тексті, з пробілами та без. Також пораховано частоту з'явлення біграм, з пробілами та без.

Нижче приведено табличку частоти букв з пробілами

	Letter	Frequency
0	Я	15113
1		122759
2	П	18364
3	И	48677
4	Ш	5761
5	у	18069
6	В	29892
7	С	37276
8	О	77949
9	Т	38333
10	н	47567
11	Л	32576
12	Ь	11865
13	Г	12497
14	Д	20506
15	е	58254
16	а	48719
17	р	30896
18	ж	6381
19	K	21599
20	ч	9628
21	Ю	4186
22	М	25307
23	Й	8574
24	б	11589
25	Ы	15802
26	Щ	3251
27	3	12064
28	Х	7634
29	Э	2266
30	ц	2209
31	ф	1135

Та табличка частоти букв без пробілів

	Letter	Frequency	
0	Я	15113	
1	П	18364	
2	И	48677	
3	Ш	5761	
4	у	18069	
5	В	29892	
6	С	37276	
7	0	77949	
8	Т	38333	
9	н	47567	
10	Л	32576	
11	Ь	11865	
12	Γ	12497	
13	Д	20506	
14	е	58254	
15	a	48719	
16	р	30896	
17	Ж	6381	
18	K	21599	
19	Ч	9628	
20	Ю	4186	
21	M	25307	
22	Й	8574	
23	б	11589	
24	Ы	15802	
25	Щ	3251	
26	3	12064	
27	Х	7634	
28	Э	2266	
29	ц	2209	
30	ф	1135	

Частота біграм з пробілами

	Diarom	Execuency	
	втаташ	Frequency	
0	яп	1072	
1	ПИ	732	
2	ИШ	543	
3	шу	190	
4	ув	862	
851	ЧХ	2	
852	чэ	1	
853	чя	1	
854	ПЖ	1	
855	гц	1	

Частота біграм без пробілів

[856	rows x	2 columns]
	Bigram	Frequency
0	яп	1072
1	ПИ	732
2	ИШ	543
3	шу	190
4	ув	862
851	ЧХ	2
852	чэ	1
853	чя	1
854	ПЖ	1
855	гц	1

Значення ентропії

```
Letter entropy with spaces: 4.4056274433859866

Letter entropy without spaces: 4.470698783872791

Bigram entropy with spaces: 3.9998311155165993

Bigram entropy without spaces: 4.1493397524316915

Bigram entropy with spaces and with intersections: 3.9993456250990387

Bigram entropy without spaces and with intersections: 4.15020687712771

Letter redundancy with spaces: 0.13402267871833318

Letter redundancy without spaces: 0.1137292848082665

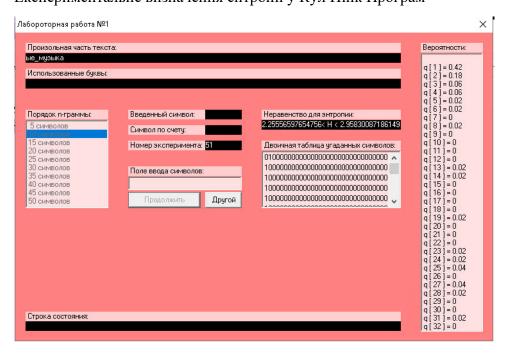
Bigram redundancy with spaces: 0.21378666727842555

Bigram redundancy without spaces: 0.177435455229774

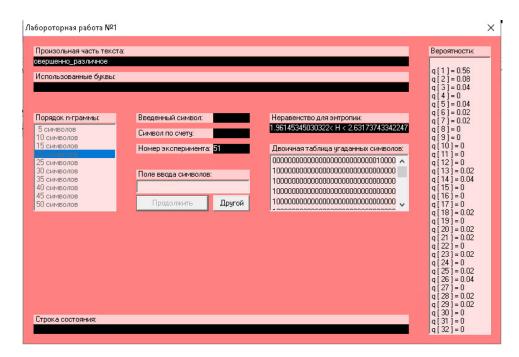
Bigram redundancy with spaces and with intersections: 0.21388209606733466

Bigram redundancy without spaces and with intersections: 0.17726355654868342
```

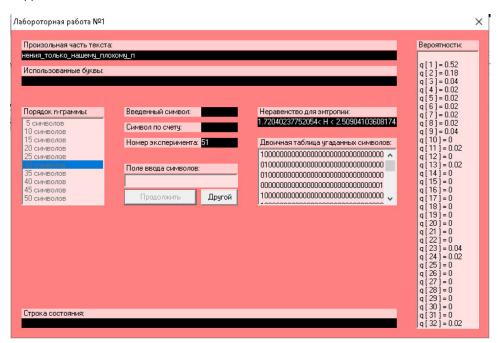
Експериментальне визначення ентропії у Кул Пінк Програм



Надлишковість == 0.418 < R < 0.557



Надлишковість == 0.418 < R < 0.482



Надлишковість == 0.506 < R < 0.661

Висновок

У ході лабораторної роботи було написано код на мові програмування Python 3 у середовищі операційної системи FEDORA 38. За допомогою коду було вирахувано частоти з'явлення літер у тексті творів Говарда Філіпса Лавкрафта росіянською мовою.

Було знайдено ентропію літер та біграм у цьому тексті. Крім того, було експериментально визначено ентропію за допомогою програми КУЛПІНКПРОГРАМ.