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

ФІЗИКО- ТЕХНІЧНИЙ ІНСТИТУТ

Кафедра інформаційної безпеки

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

з дисципліни

Криптографія

3 теми: « Експериментальна оцінка ентропії на символ джерела відкритого тексту »

Виконав студент групи ФБ-91 Свищо Максим Іванович

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:

- 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 Н1 та Н2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення Н1 та Н2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення Н1 та Н2 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:

Для тексту була взята одна з книжок про Гаррі Поттера і збережена вона у файлі text2.txt

Entropy with spaces: for letters - 4.402123312043531, for bigrams - 4.012884986309646, for bigrams with step 2 - 4.012673124538638 Entropy without spaces: for letters - 4.484090615985383, for bigrams - 4.1811174494425, for bigrams with step 2 - 4.179510576963886

Для тексту з пробілами:

Частоти літер:

H1 = 4.402123

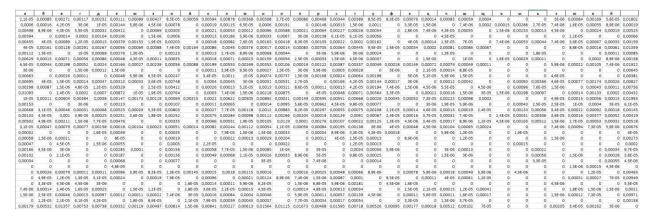
Надлишковість: 0,1347

1	Letter	Frequency	18	г	0,018794
2		0,159436	19	я	0,016434
3	0	0,092857	20	3	0,015564
4	а	0,069884	21	ь	0,015017
5	e	0,060188	22	ы	0,014766
6	и	0,059404	23	б	0,013728
7	н	0,054769	24	ч	0,012561
8	т	0,047779	25	ж	0,008366
9	С	0,045073	26	й	0,007946
10	р	0,045024	27	ш	0,007353
11	л	0,044638	28	x	0,006797
12	В	0,033669	29	ë	0,004341
13	M	0,027805	30	ю	0,004123
14	К	0,027785	31	щ	0,002555
15	Д	0,02692	32	э	0,002336
16	y	0,025665	33	ф	0,00228
17	n	0,023835	34	ц	0,002099
18	г	0,018794	35	ъ	0,000205

Частоти біграм:

H2 = 4.01288

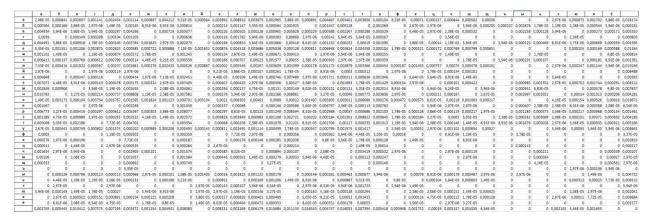
Надлишковість: 0,21122



Частоти біграм з кроком 2:

H3=4.01267

Надлишковість: 0,21126



Для тексту без пробілів:

Частоти літер:

H1=4.4840

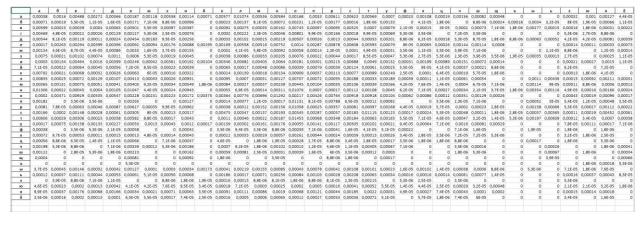
Надлишковість: 0,11109

1	Letter	Frequency			
2	0	0,11047	18	я	0,019552
3	a	0,08314	19	3	0,018516
4	e	0,071605	20	ь	0,017866
5	и	0,070672	21	ы	0,017567
6	н	0,065157	22	б	0,016332
7	т	0,056842	23	ч	0,014943
8	С	0,053622	24	ж	0,009953
9	р	0,053564	25	й	0,009453
10	л	0,053105	26	Ш	0,008748
11	В	0,040056	27	X	0,008086
12	M	0,03308	28	ë	0,005165
13	к	0,033055	29	ю	0,004905
14	д	0,032026	30	щ	0,003039
15	у	0,030533	31	Э	0,002779
16	п	0,028356	32	ф	0,002712
17	г	0,022359	33	ц	0,002497
18	я	0,019552	34	ъ	0,000244

Частоти біграм:

H2=4.18111

Надлишковість: 0,171138



Частоти біграм з кроком 2:

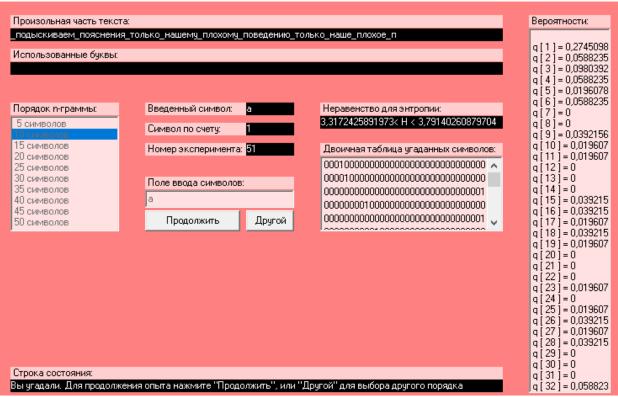
H3=4.17951

Надлишковість: 0,171455

			_	_		_		_	_	-	_					_	_		_	_		_	_				_	_			_	_	
	a	6		r	д	e	ж	3	И	Й	K	n	M	н	0	n	р	c	T	У	Ф	×	ц	ч	w	щ	ъ	ы	ь		ю	я	ë
a	0,00039																										0				0,00103		
- 6	0,00067															1,8E-05						3,9E-05				0,00023							
	0,00573																			0,00072						7,1E-06	3,5E-06					0,00034	
r	0,00465			0,00023																0,00071				7,8E-05		0	0	0			2,5E-05	-,	
A	0,00534																			0,0021												0,00078	
e	0,00016																										0				0,00011		
360	0,00121											3,5E-05				0,00014				0,00031				3,5E-05		0	0		1,4E-05			7,1E-06	
3	0,00742															0,00027						2,5E-05				7,1E-06	2,8E-05	0,00035				0,00025	
и																				0,00122							0	0			0,00025		
й				0,00038																0,00024						3,5E-06	0			7,4E-05		8,5E-05	
K	0,0078	0,00019	0,00073	0,00034	0,00025	0,00059	6E-05	0,00014	0,00336	0	0,00025	0,00142	0,00017	0,00125	0,00913	0,0004	0,00197	0,00073	0,00105	0,00258	3,5E-05	0,00013	4,6E-05	0,0002	3,9E-05	0	0	0	0	0,00011	0	4,2E-05	0
л	0,00841	0,00024	0,00077	0,00138	0,00108	0,00419	0,00045	0,00023	0,00929	0	0,00093	0,00068	0,00034	0,00129	0,00698	0,00078	0,00054	0,0028	0,00032	0,00192	0,00036	0,00011	7,1E-06	0,00058	0,00057	0	0	0,00111	0,00434	0,00014	0,00088	0,00218	0,00031
M	0,0035	0,00131	0,00083	0,00059	0,00041	0,00412	0,00011	0,00034	0,00563	3,5E-06	0,00079	0,00031	0,00036	0,00245	0,00419	0,00095	0,00033	0,00103	0,00036	0,00259	4,2E-05	8,8E-05	2,1E-05	0,0004	7,4E-05	7,1E-06	0	0,00112	9,9E-05	0,00011	1,1E-05	0,0006	0,00013
н	0,01316	0,00017	0,00046	0,00037	0,001	0,01052	3,9E-05	0,00022	0,00942	0	0,00053	7,8E-05	0,00015	0,00313	0,01053	0,00077	0,00017	0,00112	0,00105	0,00451	6E-05	7,4E-05	0,00026	0,00033	1,4E-05	3,2E-05	0	0,00344	0,00112	4,6E-05	0,00015	0,00175	0,00024
0	0,0002	0,0047	0,00941	0,00523	0,00552	0,00228	0,00229	0,00223	0,00173	0,00391	0,00363	0,00776	0,00713	0,01304	0,00212	0,00433	0,00773	0,00909	0,0095	0,00101	0,00064	0,00085	9,9E-05	0,00357	0,00134	0,00025	0	0	0	0,0004	0,00019	0,00096	0,00016
n	0,00183	0	3,5E-06	3,5E-06	0	0,00264	0	0	0,00125	0	0,00014	0,00071	1,8E-05	0,00018	0,01116	7,8E-05	0,00776	6E-05	0,00011	0,0009	0	0	7,1E-06	2,8E-05	1,1E-05	0	0	0,00053	3,2E-05	3,5E-05	0	0,00051	2,5E-05
р	0,00821	6,4E-05	0,00061	0,00046	0,00086	0,00624	0,00036	9,9E-05	0,00986	0	0,00033	0,00012	0,0015	0,00158	0,01064	0,00027	0,00356	0,00082	0,00098	0,00266	5,7E-05	0,00017	4,6E-05	0,00016	0,00025	3,2E-05	0	0,0016	0,00063	5,7E-05	0,00024	0,00107	0,00023
c	0,00166	0,00019	0,00149	0,00031	0,00063	0,00316	8,8E-05	0,00016	0,00328	0	0,0048	0,00343	0,00136	0,00198	0,00346	0,00287	0,00039	0,00184	0,01112	0,00119	0,00016	0,00022	5,7E-05	0,00049	0,00012	0	3,2E-05	0,00035	0,00362	0,00011	0,00022	0,00462	0,00064
T	0,00604	0,00028	0,00304	0,0002	0,00041	0,00571	7,8E-05	0,00017	0,00446	0	0,00121	0,00045	0,00025	0,00184	0,01441	0,00066	0,00343	0,00181	0,00058	0,0017	4,2E-05	5,7E-05	4,2E-05	0,00049	2,1E-05	1,4E-05	7,1E-06	0,00196	0,00607	9,2E-05	3,5E-05	0,00063	0,00036
y	0,00019	0,00074	0,00152	0,00135	0,00229	0,00033	0,00138	0,00041	0,00119	0,00016	0,00135	0,00297	0,00163	0,00177	0,00057	0,00141	0,00121	0,00298	0,002	0,00023	0,0001	0,00061	7,1E-06	0,00153	0,00084	0,00031	0	0	0	6,7E-05	0,00111	0,00019	1,1E-05
ф	0,0004	0	0	7,1E-06	2,8E-05	0,00056	0	0	0,00045	0	0	9,5E-05	0	7,1E-06	0,0004	1,1E-05	0,00046	2,5E-05	2,8E-05	6E-05	0,00019	0	0	1,1E-05	1,4E-05	0	0	7,1E-06	0	3,5E-06	0	3,5E-06	0
×	0,00074	8,8E-05	0,00055	9,9E-05	0,00017	0,00013	4,9E-05	0,00013	0,00039	0	0,00024	0,00035	0,00032	0,00056	0,00262	0,00042	0,00024	0,00042	0,00013	0,00027	3,2E-05	2,1E-05	7,1E-06	7,4E-05	7,8E-05	7,1E-06	0	0	0	1,8E-05	0	2,8E-05	0
ц	0,00057	1,4E-05	5,7E-05	7,1E-06	7,1E-06	0,00071	0	7,1E-06	0,00039	0	6,4E-05	0	0	1,4E-05	0,00026	2,1E-05	7,1E-06	3,5E-05	3,2E-05	0,00017	3,5E-06	3,5E-06	3,5E-06	7,1E-06	0	0	0	0,00018	0	3,5E-06	0	3,5E-06	0
ч	0,00198	1,1E-05	1,1E-05	0	7,1E-06	0,00343	0,00011	3,5E-06	0,00182	0	0,00066	8,8E-05	0	0,00104	0,00013	1,4E-05	3,5E-05	7,1E-06	0,00407	0,00071	1,4E-05	0	0	3,5E-06	0,00017	0	0	0	0,00024	0	0	0	0,00038
w	0,00123	0	4,2E-05	1,1E-05	0	0,00223	0	0	0,00161	0	0,00055	0,0008	3,2E-05	0,00034	0,00037	1,4E-05	5,7E-05	3,5E-06	0,00011	0,0003	0	0	0	3,5E-06	7,1E-06	0	0	0	0,00066	0	3,5E-06	0	0,00029
щ	0,00039	0	0	0	0	0,00079	0	0	0,00095	0	3,5E-06	0	0	4,9E-05	0	0	7,1E-06	0	0	0,00019	0	0	0	0	0	0	0	0	2,1E-05	0	0	0	0,0007
ъ	0	0	0	0	0	2,8E-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	3,5E-06	0,00014	1,1E-05
ы	3,2E-05	0,00048	0,00144	0,00031	0,00037	0,00132	0,0001	0,00031	0,00037	0,0018	0,00043	0,00215	0,00146	0,00087	0,00039	0,00078	0,00042	0,0011	0,00093	0,00025	1,8E-05	0,00105	1,8E-05	0,0004	0,00063	1,8E-05	0	7,1E-06	0	6E-05	0	7,4E-05	0
ь	0,00011	0,00039	0,00115	0,00041	0,00055	0,00059	4,9E-05	0,00057	0,00068	0	0,00192	0,00016	0,00075	0,00263	0,0008	0,00105	0,0003	0,00207	0,00072	0,00033	0,00015	0,00014	0,00013	0,00059	0,00077	1,4E-05	0	0	0	0,00019	0,00031	0,00045	8,5E-05
9	0	3,2E-05	1,1E-05	7,1E-06	1,4E-05	0	0	1,4E-05	3,5E-06	1,8E-05	0,0002	0,00013	3,5E-06	9,9E-05	3,5E-06	3,5E-06	0,0001	2,5E-05	0,0022	0	7,1E-06	2,1E-05	0	7,1E-06	0	0	0	0	0	3,5E-06	0	0	0
ю	3,9E-05	0,00022	0,00017	0,00013	0,00041	4,9E-05	3,9E-05	8,5E-05	0,0001	1,8E-05	0,00018	6E-05	0,00037	0,00028	0,00021	0,0005	0,00017	0,00041	0,00053	3,9E-05	7,1E-06	2,8E-05	2.1E-05	0,00031	4,9E-05	0,00045	0	0	0	1,8E-05	3,2E-05	3,2E-05	3,5E-06
	0.00011	0.00039	0.00178	0.00072	0.00162	0.00031	0.00022	0.00067	0.0006	3.2E-05	0.00088	0.00107	0.00084	0.00191	0.00065	0.00127	0.00044	0.00186	0.00212	0.00031	3.5E-05	0.00028	6.4E-05	0.00046	9.9E-05	0.00019	0	0	0	0.00013	0.00015	0.00017	0
ě																0.0003								7.1E-05		0	0	0		2.8E-05		2.1E-05	0
_	_	-,	-,	-,	.,	.,	.,	-,	.,	-,	-,	-,	-,	-,	-,	-,	-,	-,	-,	-,	-	-,	-,	.,	,	_	_	_	_	-,	_	-,	_

Робота з CoolPinkProgram:

Лабороторная работа №1 X



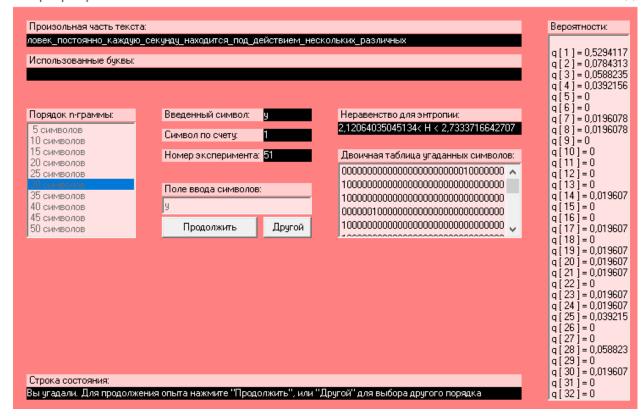
Надлишковість: 0,2417186 < R10 < 0,336551

Лабороторная работа №1

Х



Лабороторная работа №1



0,453325668 < R30 < 0,57587194

Висновки: У ході виконання цієї лабораторної роботи я виміряв частоти повторювання символів і біграм у тексті і навчився вимірювати по цим даним ентропію і надлишковість мови. Було помічено, що у тексті без пробілів ентропія більше. Також доволі корисною і цікавою була робота з наданою програмою CoolPinkProgram.