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

Експериментальна оцінка ентропії на символ джерела відкритого тексту

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.

Виконання роботи

Для обчислення усіх значень було використано текстовий файл розміром 1.4 МБ з текстом російською мовою.

Обчислення значень для тексту з пробілами

Частота літер:

H = 4.3760216803315295 R = 0.12479566393369412

	_
М	0.02525591601253788
И	0.05757881565742934
X	0.006924827618239321
a	0.072695207653342
Д	0.04421755781597693
	0.15850677070797917
б	0.013031906283999566
у	0.025292510630032638
ŗ	0.0161185215211538
К	0.030777480798381954
0	0.09311781918599496
В	0.03963197074682577
С	0.04264539674899048
Ţ	0.050744067097638665
e	0.06840659998001371
р	0.0399444324808195
ч	0.013383777606064573
ь	0.015226175848396944
П	0.023920212473979116
Я	0.01604392480087602
ж	0.007580715762568492
Ы	0.014559027821761694
Н	0.05395454104015978
ц	0.0027910433266196285
й	0.009860841929549731
3	0.01496297609949232
щ	0.007349888175293848
ф	0.0018170635071436917
Д	0.02365138278392145
ю	0.0044448385403251574
Щ	0.0029782388699582116
э	0.002585550474533665

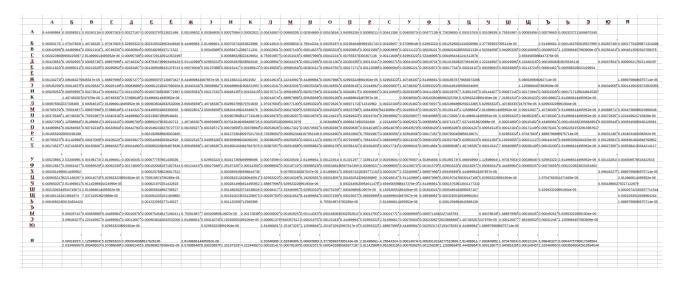
Частота біграм:

H = 3.986432684774764 R = 0.20271346304504712

	A	Б	В	Е	Л	E	Ê	ж	3	И	K	Л	M	H	0	п	P	C	y	Φ	X	П	ч	Ш	ш	ъ	ь	Э	ю	Я	
				, i								, i					, i														
١	1.407487269	0.00055595	0.00246732	0.000902191	0.002239312	0.001102062	25318443995	0.00108798	0.00387481	0.00029275	0.000577061	0.004253421	0.008643371	0.00310491	0.005126068	3.237220711	0.00086560	0.004363211	0.00403104:	0.00489805!	7.60043125	0.000187199	0.001125989	7.17818507	0.00092190	0.00079100	0.000342019	1064344688			
5	0.000054077				7.03743634	0.000010505	010700500	1.105000011	1 40740720	0.000558704	0400010000	0.00015004	0.000000731	4 50005000	0.00031668	0.00104000	0443330700	0.00107100	0.16343616	. 00046180	0.001014061	F133866113	0.01407450	. 407407200	F 60004007	7.03743634	0.0001.00400			0001707505	2021002
3			0.00011119												0.00031666																
г			2.814974538												f0.00038987f												2776493e-06	13317930-0:0	0.002610660	0.0004041/30	200315
n n															0.00192825														00019600	0.0005812923	1422114
E															0.006256280														U.000400FR	0.0003012923	******
Ê																															
K	0.00131881	2 39272835	7772004e-05	1 407487266	0.00071781	0.007959945	7272908966	9.852410884	19435466-05	0.001130212	2772299525	0.00014074	4.081713086	8 44492361	fo.00063055/	5 348451621	25505750-01	4 22246180*	2 11123090*	814974536	0.000185788	1105446497			3 51871817	11941236e-05				1.8297334500	150944
3															0.00138074												5552986e-06		00038283/	0.0001520086	
и															0.00324003								0.00144126	0 00097257							
К							2776493e-06								0.00038565																
п	0.007085290								1 54823599	0.002886756					0.000684030																
vi .															0.000458840													2776493e-010	0.00085856*	0.0035046433	900501
H	0.00358064	1.12598981	6.89668761	7.60043125/	7.03743634	0.003110546	865103605			0.002800899	665862522	0.00010837(0.000184380	8.44492361	0.001277996	0.00290505.	0.000173120	9.43016470	0.00027023	7.60043125/	0.00172557	4.222461807	832948e-06	7.03743634	2.67422581	1.125989815	2.8149745385	552986e-010	0.00069811	4.5039592616	88478
								1.12598981	5.62994907	0.007026176	448234026	0.000258977	6575470875		0.00309928/	0.009724329	54343928	3.94096435.	0.00049684	0.00081634	0.00373687	6.896687619	460482e-05	0.00030401	0.00016749	4.22246180	0.000173120	341211508	0.003194996	0.0009204966	74107
ū	1.125989815	0.00298387	0.00695439/	0.00434632	0.00450114	0.001529938	96617048049	0.00156512	0.00148208	0.000662926	0.00381006	0.00255458f	0.00627176	0.00566935	0.00537097	0.00019423	0.00135259	0.00617323	0.00557083/	0.00648288/	5.62994907	0.00030683	0.00048417	0.00012104	0.00171009	0.000772710	0.0002209755	0127659095			
P	0.001667872	4140940144				0.002123896	289339973			0.001251250	1823878304	9.00791852	0.000896569	3905298626	0.00025475	0.00917259	6.19294398	0.00649696	3.37796944	0.00015200/	0.000799452	7689497049		4.22246180	2.11123090	1.688984723	1331793e-05		0.00033075	8.7264210695	214264
C	0.007539909	6.75593889	0.000486990	0.001013396	0.000257570	0.005610244	255340711	0.00022801	4,64470798	0.005569427	1245316585	0.00039691.	0.00025616	0.00019704	0.00095709	0.007932596	6.33369271.	5.911446530	0.00015904/1	0.00092471	0.00249125	4.222461801	0.00021112	6.61519016	0.00018438	0.000447580	4.0817130809	051835e-010	0.00133570	0.0006235168	1602899
ŗ	0.00166646	7.03743634	0.00147926	2.67422581.	0.000249125	0.002950093	316405953	3.80021562	1.82973345	0.001800176	2174061136	0.00389310f	0.00246873	0.00059395	0.00086138	0.002605250	0.00161157	0.000173126	0.000919081	0.00924719.	0.00074878	2.11123090	0.00013089	7.88192870	0.00031809	0.00013371	2.8149745385	552986e-010	0.000320901	0.0034553812	3460766
v	0.00557646	2 20272025	0.00193670;	2 01/407/636	0.00015060	0.005707360	076020060		4 22246180	0.003432861	449769197	0.00072204	0.00021253/	2 01/407/53	0.00124844	0.01325140	0.16343616	0.00205430	0.000783970	34845162	0.00196907	5 62004007	1 26673854	1 60000472	0.00035750	5 62004007	5 629949077	105976-05	0.00140889	0.0045419614	1179500
D								0.00110405/							0.00049965														0.00240002	0.004541501	41,000
x			2776493e-06			0.000503880		0.00213433	0.00021000	0.000334981					1.40748726										0.00001000	0.00003240.	0.000250552		2.00647100*	1.4074872690	7776404
			0.00016608												0.00033216											2.392728357	772004e-05			7.0374363463	
			6.333692711			0.000831824				0.000282904									0496536e-0*						85552986e-0					08688981907	
îi 🗆			1.407487269			0.003791770									0.000778340												3273026e-05			0.0002547551	057301
ũ			6.474441438			0.002059153									0.00038142															0.0003363894	
ь		10025531819				0.001247033				0.001028873								2.814974538				7014663202								2.1112309039	
ь																															
ji		0.000256163	0 00083323	0.000101336	9.85241088	0.000895161	903260585	7 31893380	7 45968252	2 814974536	0.00178328	0.000201270	0.00191418	0.00123858	0.000152008	6250819861	0.000136526	0.000251946	0.00066433*	000547512	5477490057		0.00083745	1 54823599	0 00015200	0.00049965	7.037436346	88247e-06			
ã															0.00124703																
Ô.				4.222461807											1.688984723																
			1 .	1											1			1	1					1 '	1						
Я			1.68898472				7232486e-05	5.62994907	1.40748726	2776493e-06		3.65946690X	2.67422581:	6.05219525	3.940964353	9774186e-0!	1.40748726	4.644707981	0.000170309	0.000288534	8902019181	1.407487261	6.61519016	1.68898472	0.00011400	3.23722071	0.000470100	7479387349			
		2.251979630	0.000446175	9.711662157	0.000551735	9.209415977	232405e-05	0.00011963	0.00025897	2.07422581	4.705456715	0.000199865	0.000812126	0.00030542	ru.uuu547512	5477490057	5.770697807	9.20941597	0.00075863*	7.001238588	1.12598981	7.03743634	0.000147786	4.22246180	0.00015623	2.39272835	0.000370169	1510200218			

Частота біграм з кроком 2:

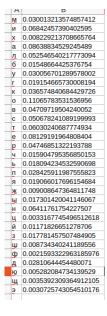
H = 3.9861208504132932 R = 0.20277582991734133



Обчислення значень для тексту без пробілів

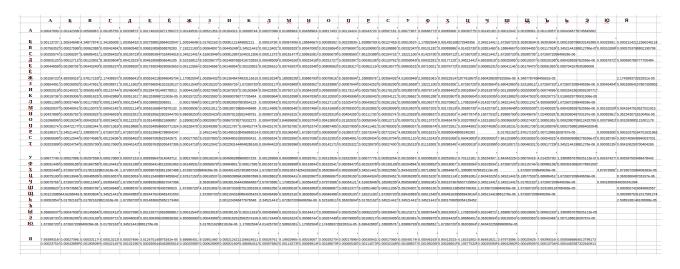
Частота літер:

H = 4.450774256302691 R = 0.10984514873946183



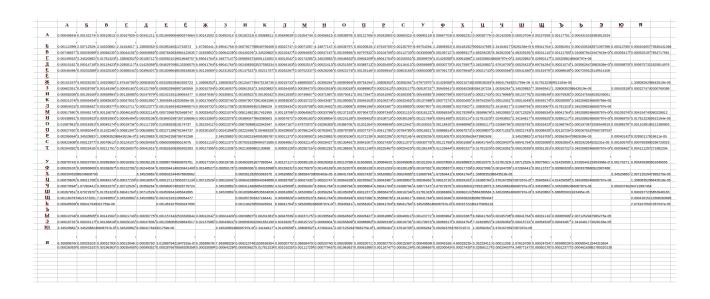
Частота біграм:

H = 4.146722577318664 R = 0.17065548453626733



Частота біграм з кроком 2:

H = 4.146564034423653 R = 0.17068719311526936



CoolPinkProgram

