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

ФІЗИКО- ТЕХНІЧНИЙ ІНСТИТУТ

Кафедра інформаційної безпеки

	U			
$KOM\Pi'$	ОТЕРНИЙ	ПРАБ	СТИКУМ	No 1

з дисципліни

Криптографія

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

Перевірила: Виконав студент групи ФБ-92

Селюх П.В Ханас Максим Любомирович

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

1.

Код програми знаходиться у файлах main.py

Результати:

H1 with whitespaces: 4.348237624893995

H2 with whitespaces: 3.935781747373973

H1 without whitespaces: 4.448697082249102

H2 without whitespaces: 4.125580893321075

Character frequency without whitespaces

Pu	
0	0.114725160004508
е	0.0871670156415899
а	0.07966000154221213
Н	0.06508609695769
И	0.06484764723677108
Т	0.06475155555819181

c 0.05293109277592251
в 0.0462616184923097
л 0.04596148028637693
p 0.04183191074150745
к 0.03302706582279983
д 0.032017510039207775
м 0.03143977365071268
y 0.02965199390233052
п 0.027441885295007386
ь 0.023226900924734117
я 0.02136438320412364
ч 0.018103197717526054
6 0.017389034871788788
г 0.016890781723599996
ы 0.016513532911399915
з 0.015397208596053123
ж 0.011408810776503806
й 0.010013701961575193
x 0.008508265663833347
ш 0.00823066748127102
ю 0.005617211087318864
э 0.0035269204989649386
щ 0.00299070520615224
ц 0.00277242287456477
ф 0.0012444465534524791
0.0

Character frequency with whitespaces

	0.17087078018969903
0	0.09512198240714967
е	0.07227271967210373
a	0.06604843492878171

Н	0.053964784791027114
И	0.053767079159957626
Т	0.053687406741466936
С	0.04388671565700729
В	0.03835685964769053
Л	0.038108006291170234
р	0.03468405951628022
K	0.027383705318281475
Д	0.02654665311907682
М	0.02606763499802786
У	0.024585334570059104
П	0.022752868944773245
Ь	0.019258102242335956
Я	0.01771383437776333
Ч	0.015009890199603998
б	0.014417756916500355
Γ	0.014004640672474556
Ы	0.013691852659140738
3	0.012766275550501985
Ж	0.009459378378085954
й	0.008302652894813719
Х	0.0070544516717929145
Ш	0.006824286907264256
Ю	0.004657393846338462
Э	0.002924272841639757
Щ	0.002479681074259612
Ц	0.0022986968149721195
ф	0.001031806999959672
	0.17087078018969903

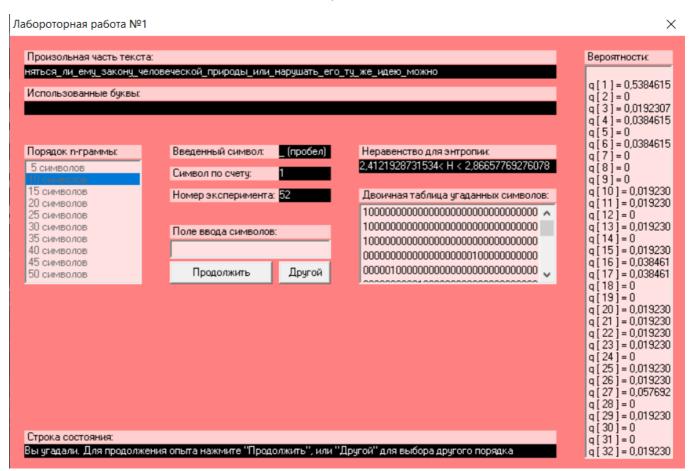
Bigram frequency with whitespaces (bigram_frec_with_whitespace.xls)

	a	б	В	Γ	Д		е	Ж	3	И	Й	K	Л	М	Н	0	П	р	С	Ţ	у ф	Х		Ц	ч	Ш	Щ	Ы	Ь	3	Ю	Я	
a	2.9508303	14 0.000491	805 0.00259	8697 0.000	640330 0.0	01813777	0.001177381	1 0.0016091	86 0.0033629	962 0.000111	147 0.00088328	1 0.00549739	6 0.00705	5435 0.00337	2799 0.0030865	8 1.0819711	15 0.00074754	3 0.00169672	7 0.00457	78705 0.004208867	6.295104670 5.	016411534 0.0	00890167	4.1311624	40 0.0009098	39 0.00088131	4 0.0002390	17 0.0	0.0	2.9508303	14 0.0008508	22 0.0019357	44 0.017237767
б	0.0006108	21 9.836101	048 4.81968	9513 0.0	1.9	67220209	0.002112794	43.9344404	19 0.0	0.000789	838 0.0	0.00013475	4 0.00069	395 4.22952	3450 0.0002518	04 0.0019170	156 0.0	0.00115574	10.00018	39836 3.934440419	0.001273775 0.	0 4.6	22967492	9.8361010	48 1.9672202	09 1.08197111	5 0.0001632	79 0.0036826	36 0.00018393	35 0.0	7.8688808	38 0.0006059	03 0.000444591
В	0.0056429	71 1.967220	209 2.75410	3293 1.377	7054146 0.0	00840986	0.004955427	70.0	0.0005360	067 0.003128	863 0.0	0.00010819	7 0.00051	7378 0.00010	0328 0.0015334	18 0.0061465	79 0.00037868	9 0.00049869	0.00359	98045 0.000192787	0.000581313 0.	0 5.9	01660628	31.5737761	67 0.0001908	20 0.00061869	90 9.8361010	48 0.0026754	19 0.00015737	77 0.0	0.0	0.0001859	02 0.005649856
Γ	0.0009865	60 0.0	5.90166	0628 0.0	0.0	01099676	0.000179017	70.0	9.8361010	048 0.000506	559 0.0	4.81968951	30.00126	3.83607	9408 0.0002645	91 0.0075305	18 0.0	0.00051049	3 1.67213	37178 5.901660628	0.000446558 0.	0.0		0.0	3.4426353	66 9.83610104	18 0.0	0.0	0.0	0.0	0.0	0.0	0.001062298
Д	0.0051285	43 2.754108	293 0.00089	134 7.868	880838 3.7	37718398	0.004217720	01.9672202	09 1.3770541	146 0.002303	614 0.0	0.00021737	7 0.00056	9510 7.081992	2754 0.0017990	22 0.0036009	96 8.16396387	0 0.00184721	90.00026	32623 0.000268525	0.001957384 0.	0 4.5	24606482	0.0002645	91 5.4098555	76 6.68854871	2 1.9672202	09 0.0004249	119 0.00105836	640.0	1.0819711	15 0.0004259	03 0.000870494
е	8.8524909	43 0.001318	037 0.00106	3200 0.003	269519 0.0	02673452	0.001731153	30.0007859	04 0.0011773	381 9.639379	027 0.00210886	0 0.00110852	28 0.00551	3134 0.00418	3293 0.0064023	18 0.0002242	63 0.00118328	2 0.00578854	5 0.00424	11326 0.005434445	0.0001554108.	852490943 0.0	100462296	0.0003265	58 0.0011704	96 0.00101213	34 0.0009393	47 0.0	0.0	0.0	0.0002291	81 0.0002144	27 0.019436135
ж	0.0012098	40 2.360664	251 0.0	3.934	440419 0.0	00698363	0.004381983	31.0819711	15 0.0	0.001248	201 0.0	0.00010032	81.77049	3188 3.93444	0419 0.0007032	31 1.6721371	78 0.0	4.91805052	4 6.88527	70733 0.0	0.000260656 0.	0.0		9.8361010	48 3.9344404	19 0.0	0.0	0.0	4.91805052	240.0	0.0	0.0	0.000714100
3	0.0044636	22 0.000115	082 0.00085	8691 0.000	341312 0.0	00685576	0.000173115	54.6229674	92 8.8524909	943 0.000348	197 0.0	0.00011016	64 0.00023	3115 0.00018	7869 0.0017606	62 0.0004859	03 0.0	0.00023311	51.18033	32125 7.868880838	0.000607871 0.	0.0		0.0	1.8688591	99 8.85249094	13 0.0	0.0002567	722 0.00014163	39 0.0	1.9672202	09 0.0004209	985 0.001238365
И	6.9836317	44 0.000651	149 0.00242	648 0.000	1525247 0.0	01686891	0.001655415	5 0.0002095	08 0.0016032	284 0.000516	395 0.00094229	8 0.00229181	10.00483	4443 0.00246	1926 0.0030068	96 0.0001937	71 0.00021049	2 0.00080459	3 0.00185	57055 0.004116408	8.852490943 1.	377054146 0.0	01666235	0.0008095	11 0.0014891	85 0.00062164	1 0.0001278	69 0.0	0.0	9.8361010	48 0.0003000	01 0.0010898	339 0.017577112
Й	0.0	0.0	0.0	0.0	0.0	00272459	0.0	0.0	1.9672202	209 0.0	0.0	0.00012393	34 0.00020	3525 4.22952	3450 0.0003196	73 1.9672202	09 0.0	0.0	0.00041	12132 0.000359017	0.0 0.	0.0	1	3.0491913	24 0.0001770	49 0.00018491	18 9.8361010	48 0.0	0.0	0.0	0.0	0.0	0.006167235
K	0.0068242	86 0.0	0.00024	984 0.0	1.9	67220209	0.000374755	5 3.9344404	19 1.0819711	115 0.001958	367 0.0	1.08197111	50.00050	3559 2.95083	0314 0.0005252	47 0.0085564	24 0.0	0.00144689	0.00019	98689 0.000476067	0.001179348 0.	0 3.4	42635366	4.9180505	24 0.0	1.08197111	15 0.0	0.0	0.0	0.0	0.0	0.0	0.005015427
Л	0.0053891	99 2.557386	272 1.37705	146 0.000	128852 1.9	67220209	0.003665914	40.0003059	02 4.9180505	524 0.005320	347 0.0	0.00026262	34.52460	6482 9.83610	1048 0.0002468	36 0.0052691	99 5.70493860	7 0.0	0.00139	97709 0.000101311	0.0014567261.	967220209 0.0		0.0	0.0002104	92 0.0	9.8361010	48 0.0005960	067 0.00463673	38 0.0	0.0007357	40 0.0012334	147 0.006971828
М	0.0028731	25 1.180332	125 9.83610	048 7.672	1588170.0		0.003890177	70.0	0.0	0.002616	402 0.0	7.08199275	48.95085	1953 2.36066	1251 0.0015423	00 0.0036531	27 3.73771839	8 6.29510467	0 8.65576	38922 2.852469303	0.002119679 1.	081971115 0.0	1	8.8524909	43 4.9180505	24 5.90166062	28 0.0	0.0008016	.42 7.86888083	88 9.8361010	48 3.9344404	19 0.0003383	861 0.007585601
Н	0.0099826	58 9.836101	048 1.08197	115 2.557	386272 0.0	00279345	0.010064298	8 1.9672202	09 1.2786931	136 0.007495	108 0.0	0.00024983	36 0.0	0.0	0.0023921	39 0.0095970	183 0.0	5.11477254	5 0.00024	11968 0.000363935	0.002778698 3.	442635366 3.9	34440419	0.0001445	90 0.0001888	53 3.93444041	198.3606858	90 0.0024708	.28 0.00119311	190.0	0.0002085	25 0.0017813	317 0.004294441
0	5.9016606	28 0.003420	012 0.00771	6421 0.004	076080 0.0	04831492	0.001681973	3 0.0020724	66 0.0008872	216 0.000614	756 0.00280328	8 0.00154721	8 0.00574	1332 0.00507 ⁻	1493 0.0063787	11 0.0002114	76 0.00135541	4 0.00484329	6 0.00553	32806 0.006461334	8.852490943 0.	000168197 0.0	100464263	8.3606858	90 0.0022475	49 0.00098656	60 0.0002252	46 0.0	0.0	9.8361010	48 0.0005744	28 0.0006108	321 0.024419104
П	0.0007888	55 0.0	0.0	0.0	0.0		0.002820993	3 0.0	0.0	0.000786	0.0 888	7.57379780	7 0.00062	0657 0.0	9.5410180	16 0.0090275	73 2.85246930	3 0.00664035	1 2.95083	30314 6.885270733	0.000750494 0.	0.0		8.8524909	43 1.9672202	09 6.88527073	3 0.0	0.0002163	394 0.00015245	59 0.0	0.0	0.0005724	61 4.229523450
p	0.0072619	93 0.000109	180 0.00036	0.000 0.000	151475 0.0	00300984	0.005060673	3 0.0002331	15 5.1147725	545 0.005192	477 0.0	0.00019475	46.59018	7702 0.00028	2296 0.0006373	79 0.0070800	25 9.83610104	8 9.83610104	80.00011	11147 0.000616723	0.002522959 0.	000286230 6.0	100021639	1.6721371	78 7.1803537	65 0.00021737	77 1.2786931	36 0.0008075	43 0.00095213	34 0.0	0.0001986	89 0.0009619	70 0.000757379
C	0.0016790	22 8.754129	932 0.00133	726 2.950	1830314 0.0	00201640	0.004417392	2 2.3606642	51 9.8361010	048 0.001389	841 0.0	0.00346624	2 0.00264	9845 0.00109	9676 0.0008754	12 0.0027019	76 0.00169180	9 0.00015049	20.00060	0.009360033	0.0006895102.	950830314 0.0	100183935	5.1147725	45 0.0003659	02 7.27871477	75 9.8361010	48 0.0001977	05 0.00306394	15 0.0	0.0002478	69 0.0035547	766 0.003679685
T	0.0056508	40 7.868880	838 0.00205	3695 9.836	101048 8.3	60685890	0.005817070	0.00	1.0819711	115 0.003232	142 0.0	0.00033639	4 0.00019	7705 2.55738	3272 0.0010868	39 0.0146184	13 6.68854871	2 0.00299115	8 0.00102	22954 6.688548712	0.001626891 0.	0 1.3	77054146	6.0983826	49 0.0003236	07 9.83610104	18 2.5573862	72 0.0013829	55 0.00658428	36 9.8361010	48 7.9672418	49 0.0003904	193 0.005895758
у	2.7541082	93 0.000720	002 0.00064	182 0.001	435087 0.0	01949515	0.000184918	8 0.0017508	25 0.0002183	361 6.885270	733 0.00022032	8 0.00063639	95 0.00145	4759 0.00156	890 0.0005075	12 3.9344404	19 0.00064524	8 0.00041311	6 0.00093	38364 0.001480333	9.836101048 1.	377054146 0.0	00386558	34.9180505	24 0.0007868	88 0.00056852	26 0.0002754	10 0.0	0.0	9.8361010	48 0.0007180	35 5.6065775	597 0.006963959
ф	0.0002203	28 0.0	0.0	0.0	0.0		4.721328503	3 0.0	0.0	0.000292	132 0.0	1.96722020	97.86888	0838 3.93444	0.0 0.419	0.0001032	79 0.0	7.67215881	79.83610)1048 1.278693136	0.000106229 1.	081971115 0.0		0.0	9.8361010	48 9.83610104	18 0.0	2.1639422	230 0.00010524	16 0.0	0.0	0.0	1.868859199
X	0.0004042	63 0.0	0.00018	885 0.0	0.0		0.000316722	20.0	7.8688808	338 0.000520	329 0.0	9.83610104	8 0.00013	3689 6.49182	6691 0.0001475	41 0.0022347	62 0.0	7.77051982	8 1.96722	20209 5.901660628	0.000145574 0.	0.0		0.0	0.0	1.47541515	7 0.0	0.0	0.0	1.9672202	09 0.0	0.0	0.002765911
Ц	0.0005075	42 0.0	2.55738	3272 0.0	0.0		0.000737707	70.0	0.0	0.000118	033 0.0	1.77049818	80.0	2.95083	0314 0.0	0.0001908	20 0.0	0.0	0.0	2.950830314	0.000181967 0.	0.0		0.0	0.0	0.0	0.0	0.0001573	77 0.0	0.0	0.0	0.0	0.000356066
Ч	0.0025013	20 0.0	9.83610		0.0		0.003809521		0.0	0.001532	464 0.0	0.00050754	21.27869	3136 9.83610	1048 0.0008291	33 7.8688808	38 0.0	8.36068589	0.00	0.004143949	0.000667871 0.	0.0		0.0	0.0	0.00013967	72 0.0	0.0	0.00024098		0.0	0.0	0.000460329
Ш	0.0008901	67 0.0	5.90166	0628 0.0	0.0		0.002113778		0.0	0.001343		0.00044655	8 0.00052	0329 1.18033	2125 0.0003954	11 0.0001750	182 0.0	0.0	0.0	4.229523450	0.000196722 0.			0.0	0.0	0.0	0.0	0.0	0.00062065		0.0	0.0	6.196743660
Щ	0.0003609		0.0	0.0	0.0		0.001350496		0.0	0.000525		0.0	0.0	0.0	4.8196895	13 0.0	0.0	2.95083031	40.0	0.0	0.000148525 0.	0.0		0.0	0.0	0.0	0.0	0.0	3.83607940	0.080	0.0	0.0	4.918050524
Ы	0.0	0.000147	541 0.00076	330 0.000	109180 0.0	00157377	0.000703281	1 2.2623032	41 4.7213285	503 5.901660	628 0.00123443	0 0.00014754	10.00187		0658 0.0001642			6 0.00019475	4 0.00060)7871 0.000707215	1.967220209 0.	0.0	100780986	9.8361010	48 0.0001711	48 0.00044950	9 6.8852707	33 0.0	0.0	0.0	0.0	2.9508303	314 0.004424278
Ь	0.0	7.081992	754 2.45902	262 0.000	127869 1.8	68859199	0.000435739	90.0	0.0001445	590 0.000111	147 0.0	0.00127574	20.0	0.00033	3361 0.0020665	64 4.9180505	24 0.0	0.0	0.00083	31150 0.000179017	0.0 1.	278693136 0.0	100125902	26.3934656	81 5.8032996	18 0.00029016	64 2.7541082	93 0.0	0.0	0.0	0.0003167	22 0.0006167	23 0.012117092
	0.0	0.0			660628 5.9			0.0	0.0	0.0)209 2.3606642			4 9.83610104	8 6.88527	70733 0.002735419	0.0 8.	852490943 4.3			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.360664251
	0.0		951 2.95083								314 9.83610104						0.0			34918 0.000307869			01660628			37 6.19674366			0.0	0.0	3.6393573		0.003046240
	0.0										157 8.95085195									34428 0.001570825						74 1.08197111			0.0	0.0			11 0.011568238
	0.0025564	02 0.006867	565 0.01679	3093 0.002	963617 0.0	08098061	0.004118375	5 0.0022524	67 0.0044518	319 0.011638	074 6.88527073	3 0.00817281	6 0.00244	0336 0.00582	1939 0.0162797	30 0.0114865	98 0.01601022	1 0.00497509	9 0.01626	61042 0.010357414	0.004567885 0.	000408198 0.0	01624923	0.0002901	64 0.0062046	12 0.00056754	13 6.5901877	02 0.0	9.83610104	18 0.0029055	84 4.5246064	82 0.0026301	73 0.0

Bigram frequency without whitespaces (bigram_frec_without_whitespace.xls)

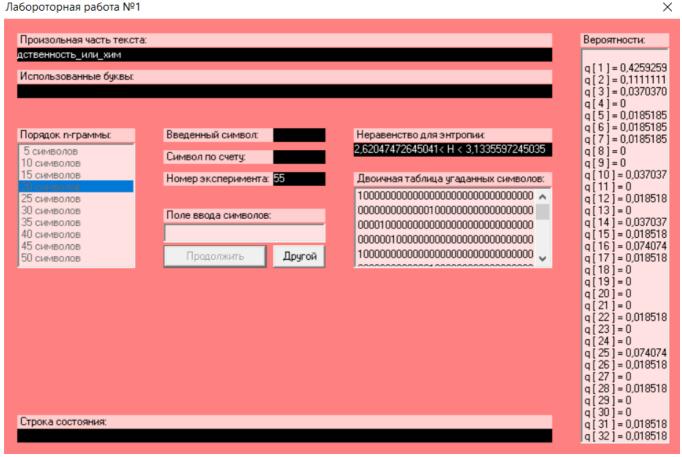
	_	-			-	-			1	-		X.			_			-		-	-			-		-											-	
	0.00020042	20 0 00420	2240000	0547007	0.004440	2020.00	24600200	00000407	* n nnn4220	3	225200046	C045700	0040052420	007002775	0 00003343	M 0 0 00 40 4	DE 4C O O	057574000	0042764	27 0 002000	200400	02502200	0074004	04.0.006344	y 1460 0 000	φ 202002400	0100000	0.004.050000	U 0 40550742	40.004000	145 0 0011329	щ	E1000	0.0	0.000346	404 0 0010285	я	62500
	0.00030013								1 4 7452680						0.00083213																779 1 3049487			0.0				
																															779 1.3049487. 161 0.0007841							
	0.00000007																											931585097			203 1 1863170		0.00322	0.000108		155 1 1863170		
_									2 2 8471608																						155 8 5414825		0.0	0.0				
																															887 0 0013002i			2400 0.00 1270		663 0 0002823		
	0.00028471														3 32168765													1 186317019				0.001130	0.0	U.U E 024E0E	0.000462		2 9657925	
	0.00146626																														607 1 5422121:	0.0	0.0					
																															862 0 0008197			0.000170		381 0 0003665		
	9.72779955																														500 0.0002752			0.0		348 8.3042191		
																															400 2.9657925			0.0		348 8.3042191 768 1 1863170		
	0.00828761																														400 2.9657925 219 2.6098974							
																															788 4 8638997							
																															788 4.8638997 876 3 0844242							
																															617 0 0012634					399 0.0002514		
-	0.00025743										1058 0 0002				0.00735635																442 8 3042191			0989 0 000183		0.0000967	0.0001326	
																															327 0 0002657					155 0 0000400		
P									2 0 0002645																						960 0 0001055							
-	0.00209740																														181 2 9657925							
	0.00702336																														070 0.0007058			0.007941		700 0.0008671		
	0.00023726					0.00		.69432169			7019 0 0003				1.06768531											0128122 1.30			1.30494672		038 1.1863170			7442 0.000126		0.0000671	0.0002340	0.0
	0.00020032					0.0																							0.0		857 4 1521095			0.000120	6 168848	E01 0.0	4.0334778	
	0.00053026																											1 186317019			529 1 1863170		0.00018			116 1 1863170		
	0.00302273														2.49126574													3 304219136			721 0 0001684		0.00010		647 5.931585		1 1863170	
7									1 1 1863170						0.00062993																019 1 1863170		0.0	0.000230		0.0	2.3726340	
ш	0.00107480		0.00	23300423	0.0	0.00 4.74		00163118		0.0		6346790		000043333	0.00062993	0.0		12953395 0		0.0		45268077		0.0		0179133 0 0		0.0	1.10031701	0.0	0.0	0.0	0.0	4 626636		0.0	0.0	0.0
щ			0.0	01440670	0.0	164 0 00				26 0 00022				000200700	0.0	0.0				0.0				60 0 001246					0.0 E 021E9E00	7.0.000247	590 0 0005516:	7 1 067696		0.0		395 1 1863170	10 5 6042216	602 0 0
									0 0.0001091																						547 0 00033772			0.0		263 0.0003843		
0	0.00030013	0.0000			7 117902				0.000109	0.00052	0.0014		542212125.5											55 0 003299				5 219794885		2 372634		0.0	0.0	0.0	1 542212		0.0010380	0.0
in .	8 66011424								5.8 1855874	34 9 37190	4454 0 0002																				020 9 1346410	0.0	860 0.0	0.0		587 4 5080046	74 6 2874803	
																															306 5 5756899			0.0		810 0.0001138		
л	0.00033333	0.00048	0.0	01134091	0.000281	0.00	12040130	.00045186	0.0002868	0.00000	0.0011	0.04720.	0001079540	n	0.00116021	0.00076	0.0	021121400.	.0000041	0.001298	0.0		1.0020641	0.002/30	0.000	0.0	13033932	0.000341039	0.00011131	0.000032	0.0	0.000144	0.0	0.0	0.000189	0.0001136	0.0002590	0.0
	0.0	0.0	0.0		0.0	0.0		.0	0.0	0.0	0.0	0.	0 0	.0	0.0	0.0	0.0	0.	.0	0.0	0.0		1.0	0.0	0.0	0.0		7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

H10



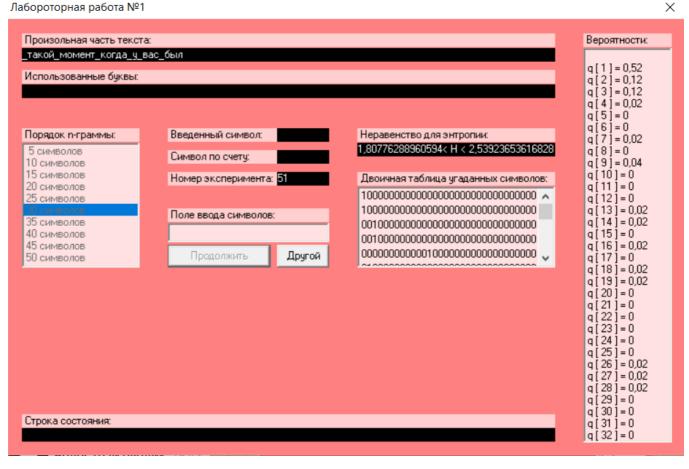
H20

Лабороторная работа №1



H30

Лабороторная работа №1



Оцінка надлишковості мови:

Н	R
H1 with whitespaces	0,12618
H1 without whitespaces	0,114166
H2 with whitespaces	0,44269569
H2 without whitespaces	0,368318
H10	0,517>R>0,427
H20	0,475>R>0,373
H30	0,638>R>0,492

Висновки:

Під час виконання даної лабораторної роботи я навчився вимірювати частоти символів та біграм у тексті, визначати ентропію. Також, я навчився визначати надлишковість мови.