



МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

Кафедра інформаційної безпеки

Лабораторна робота №1

з дисципліни «Криптографія» на тему:

«Експериментальна оцінка ентропії на символ джерела
відкритого тексту»

Перевірив:

Виконали:

Студентки групи ФБ-91

Легенчук М. О.

Осьмак А. А

Київ – 2021

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

H(1)

Буква	Повторения	Частота
О	96687	0.11474
Е	73451	0.087164
А	67120	0.079651
Н	54840	0.065078
И	54646	0.064848
Т	54556	0.064741
С	44604	0.052931
В	38986	0.046264
Л	38737	0.045969
Р	35250	0.041831
К	27837	0.033034
Д	26984	0.032022
М	26496	0.031443
У	24994	0.02966
П	23123	0.02744
Ь	19569	0.023222
Я	18002	0.021363
Ч	15249	0.018096
Б	14655	0.017391
Г	14233	0.01689
Ы	13912	0.016509
З	12975	0.015397
Ж	9615	0.01141
Й	8438	0.010013
Х	7171	0.0085098
Ш	6938	0.0082333
Ю	4733	0.0056166
Э	2971	0.0035257
Щ	2521	0.0029917
Ц	2336	0.0027721
Ф	1049	0.0012448

Энтропия: 4.4487

Буква	Повторения	Частота
—	170412	0.16821
О	96687	0.095438
Е	73451	0.072502
А	67120	0.066253
Н	54840	0.054131
И	54646	0.05394
Т	54556	0.053851
С	44604	0.044028
В	38986	0.038482
Л	38737	0.038236
Р	35250	0.034795
К	27837	0.027477
Д	26984	0.026635
М	26496	0.026154
У	24994	0.024671
П	23123	0.022824
Ь	19569	0.019316
Я	18002	0.017769
Ч	15249	0.015052
Б	14655	0.014466
Г	14233	0.014049
Ы	13912	0.013732
З	12975	0.012807
Ж	9615	0.0094908
Й	8438	0.008329
Х	7171	0.0070783
Ш	6938	0.0068484
Ю	4733	0.0046718
Э	2971	0.0029326
Щ	2521	0.0024884
Ц	2336	0.0023058
Ф	1049	0.0010354

Энтропия: 4.354

без проби́на																															
А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ш	Щ	Ы	Ь	Э	Ю	Я		
0.00031447	0.0012828	0.0051775	0.0011107	0.0031708	0.0020043	0.0021337	0.0045711	0.0016839	0.0010657	0.0076838	0.0088338	0.0048429	0.0057531	0.0013742	0.0028018	0.0025953	0.0047462	0.0035512	0.0006303	0.00010087	0.0012591	8.1882e-05	0.0018607	0.0011333	0.0030261	-	-	-	-		
0.00074168	4.7468e-06	8.4255e-05	-	2.6107e-05	0.0025644	4.7468e-06	1.424e-05	0.00099208	-	0.00017563	0.00084018	5.4588e-05	0.0003299	0.002409	1.3054e-05	0.0013967	0.00025989	2.7294e-05	0.0015866	-	5.6961e-05	1.1867e-06	4.8654e-05	1.3054e-05	0.00019699	0.004443	0.00021291	0.00012026	9.4936e-06	0.00077254	
0.0068852	0.00021598	0.00048061	0.00027175	0.0013374	0.0061483	4.8654e-05	0.00079034	0.00040419	3.5601e-06	0.00064912	0.00027507	0.00025514	0.00024802	0.0078417	0.001106	0.00083187	0.0051265	0.00076304	0.00090663	1.068e-05	0.00011392	3.3227e-05	0.00040229	0.0007844	4.7468e-06	0.0032278	0.00018868	0.00032041	4.7468e-06	0.00026463	
0.0012092	2.8481e-05	0.00014478	2.136e-05	0.0013837	0.00023615	9.4936e-06	6.4081e-05	0.0006776	-	0.00011392	0.0015403	7.5948e-05	0.00043552	0.0092277	0.000178	0.00068235	0.0001424	4.7468e-05	0.00057673	3.5601e-06	5.9335e-06	-	6.1708e-05	1.1867e-06	-	-	-	9.4936e-06	1.1867e-05	-	
0.0062052	6.6455e-05	0.0011772	2.2547e-05	9.9682e-05	0.0051147	2.8481e-05	6.0521e-05	0.0028457	-	0.00034651	0.00071795	0.00010799	0.0032371	0.0043955	0.00018038	0.00022618	0.00042009	0.00035719	0.00023793	2.3734e-06	5.6961e-05	0.00032041	9.4936e-05	8.5442e-05	2.3734e-06	0.00051147	0.0012769	1.8987e-05	1.3054e-05	0.00052452	
0.00028481	0.0026119	0.0037464	0.0043706	0.00043552	0.0025205	0.0011831	0.0022381	0.001627	0.0025443	0.0022096	0.0007068	0.0060747	0.0009146	0.0017041	0.0030903	0.00076423	0.0057541	0.00079461	0.00080695	6.0521e-05	0.00094342	0.00042365	0.0020565	0.0013006	0.0011392	-	-	0.00046281	0.00028243	0.00052808	
0.0014668	6.5268e-05	5.9335e-05	1.068e-05	0.00087103	0.0052986	1.5427e-05	1.6614e-05	0.0015831	-	0.00021479	3.3227e-05	2.2547e-05	0.00095529	0.0001068	5.9335e-05	1.8987e-05	6.6455e-05	5.4588e-05	0.00032753	-	1.1867e-06	1.1867e-06	2.6107e-05	-	-	-	-	5.9335e-05	4.6281e-05	2.9667e-05	
0.0053995	0.00016495	0.0011843	0.0004462	0.00092681	0.00023853	6.6455e-05	6.1708e-05	0.0004162	-	0.00024565	0.00030261	0.00030973	0.002263	0.00066218	0.00017919	0.00030973	0.00014834	0.00012342	0.00077254	3.5601e-06	8.3069e-06	9.4936e-06	6.5268e-05	1.5427e-05	3.5601e-06	0.00030973	0.00017088	2.6107e-05	2.3734e-06	0.00051621	
0.00033346	0.0016709	0.005136	0.00099564	0.0013305	0.0024458	0.00046281	0.0025419	0.00047611	0.0011369	0.0038461	0.0060901	0.0036099	0.0057614	0.016178	0.00023805	0.0015664	0.0045047	0.0061637	0.00063013	6.8828e-05	0.0002298	0.0010217	0.00024565	0.00082001	0.00015902	-	-	0.00024446	0.00036669	0.0016353	
9.7309e-05	0.00023971	0.00057436	0.0002409	0.00072982	9.8496e-05	0.00010799	0.00015783	0.00061589	-	0.00053876	0.0003477	0.00038686	0.00098614	0.000675	0.00074762	0.00039201	0.0012567	0.00082357	0.00016258	2.9667e-05	6.6455e-05	5.9335e-05	0.00056368	0.00027531	3.5601e-06	-	-	9.0189e-05	8.3069e-06	7.7135e-05	
0.0002867	0.00062183	0.00800102	9.2562e-05	0.00620701	0.00053995	0.00014359	0.00015918	0.002784	1.1867e-06	0.00026226	0.00071202	0.00020886	0.0012104	0.010679	0.00039398	0.0019201	0.00077847	0.00101716	0.0015593	9.4936e-06	8.4255e-05	0.00022547	2.9667e-05	3.5601e-06	-	-	-	0.0001068	1.1867e-06	0.00011986	
0.00606704	0.00028599	0.00077016	0.00030142	0.00033465	0.0048144	0.00039873	0.00020411	0.00070347	-	0.0006954	0.00015664	0.00016732	0.0010395	0.0074869	0.00072388	0.00040415	0.0023876	0.00046044	0.0018975	2.3734e-05	3.6788e-05	4.7468e-06	0.00065743	2.6107e-05	3.5601e-06	0.00071914	0.00055941	0.00011155	0.00088646	0.0016341	
0.0035826	0.00029549	0.00084374	0.00028125	0.0004723	0.0048453	0.00016139	0.00023853	0.0038401	1.1867e-06	0.0004723	0.00021954	0.00027413	0.0027068	0.0050755	0.0009719	0.00034295	0.010704	0.00052808	0.0028054	4.9841e-05	8.0695e-05	1.8987e-05	0.00039992	4.8654e-05	3.5601e-06	0.00096597	9.4936e-05	0.00013054	8.3069e-06	0.00056724	
0.012102	0.00027531	0.00063488	0.00010799	0.00052333	0.012241	2.6107e-05	0.00015427	0.0093547	1.1867e-06	0.00053995	6.2895e-05	0.00013291	0.003337	0.011899	0.00057317	0.00030142	0.00086391	0.00070134	0.00035316	4.6281e-05	5.5775e-05	0.00018275	0.000375	3.0854e-05	0.00010443	0.0029739	0.00014383	5.8148e-05	0.00025158	0.0021978	
0.00025751	0.0056499	0.01259	0.0053971	0.0071439	0.0028386	0.0030795	0.00018133	0.002333	0.0033821	0.0030866	0.0073599	0.00071866	0.010379	0.023031	0.00452331	0.0065399	0.0093452	0.0097487	0.00097546	0.00027175	0.00081289	0.00015071	0.0038036	0.001265	0.00028006	-	-	1.1867e-06	0.00044026	0.00069896	
0.00095173	4.7468e-06	2.3734e-06	1.1867e-06	2.3734e-06	0.0034023	-	3.5601e-06	0.00095529	-	9.2562e-05	0.00074762	1.1867e-06	0.00011511	0.010891	4.5094e-05	0.0080102	3.9161e-05	8.5442e-05	0.00090663	-	2.3734e-06	1.068e-05	2.6107e-05	8.3069e-06	-	-	0.00026107	0.00018394	-	-	0.00068947
0.000778	0.00016851	0.00050197	0.00019936	0.00039754	0.0061138	0.00028481	7.5948e-05	0.0063512	1.1867e-06	0.00027531	8.1882e-05	0.00035007	0.00084967	0.0086	0.00031685	4.5094e-05	0.00020648	0.00078559	0.0030617	0.00034889	7.9509e-05	2.0174e-05	0.00011036	0.00026582	1.5427e-05	0.00097428	0.0011487	9.4936e-06	0.0002409	0.0011725	
0.0020981	0.00026226	0.0019747	9.8496e-05	0.00047112	0.0053983	0.00026463	9.7309e-05	0.0018619	-	0.00044287	0.0023919	0.0015178	0.00157	0.0034996	0.0024683	0.00026701	0.0010965	0.011619	0.0010241	1.1867e-05	0.00026226	6.7642e-05	0.00054113	0.00010562	3.5601e-06	0.00023853	0.0036977	7.2388e-05	0.00030142	0.0043196	
0.0070229	0.00042484	0.0031352	0.00010918	0.00045569	0.0072187	0.00013172	0.00016732	0.0043771	-	0.0007488	0.00032041	0.0026938	0.0019426	0.018083	0.00065031	0.0037566	0.0018548	0.00047705	0.00021562	2.4921e-05	5.9335e-05	7.815e-05	0.0006954	2.9667e-05	3.2041e-05	0.0016685	0.0079331	0.0001424	9.7309e-05	0.00058029	
У	0.00023734	0.0010704	0.0015914	0.0018441	0.0027543	0.00035601	0.0022464	0.00046637	0.00067998	0.00026582	0.0011558	0.0018465	0.0022623	0.0015581	0.00047112	0.0014952	0.00069778	0.0018762	0.0023307	0.00016139	3.9161e-05	0.00055063	1.3054e-05	0.0014335	0.00070608	0.0003465	-	-	0.00011274	0.00086747	0.00023497
0.00027601	1.1867e-06	1.1867e-06	-	-	5.6961e-05	-	1.1867e-06	0.00035719	-	3.5601e-06	1.068e-05	4.7468e-06	-	-	0.0001246	7.1202e-06	9.2562e-05	2.3734e-06	1.6614e-05	0.00012816	1.3054e-05	-	-	2.3734e-06	1.1867e-06	-	-	-	2.6107e-05	0.00012698	-
Х	0.0052927	0.0001341	0.00051147	6.5268e-05	0.00021717	0.00045094	4.1534e-05	8.0695e-05	0.00089477	-	0.00016258	0.00024029	0.00018631	0.00043908	0.0028932	0.0004011	0.00018394	0.00033702	0.00020055	0.00026938	7.1202e-06	1.068e-05	9.4936e-06	9.6122e-05	4.1534e-05	2.3734e-06	-	-	6.1708e-05	-	4.0348e-05
0.00062301	7.1202e-06	7.0015e-05	5.9335e-06	1.3054e-05	0.00090545	2.3734e-06	9.4936e-06	0.00017444	-	4.0348e-05	4.7468e-06	1.3054e-05	3.6788e-05	0.00026701	6.1708e-05	1.3054e-05	3.0854e-05	3.3227e-05	0.00023022	3.5601e-06	1.1867e-06	-	1.78e-05	1.1867e-06	-	-	0.00018987	-	7.1202e-06	1.1867e-06	8.3069e-06
0.0030154	2.136e-05	6.7642e-05	8.3069e-06	2.7294e-05	0.0046008	1.1867e-06	9.4936e-06	0.0018868	-	0.0006515	2.4921e-05	1.424e-05	0.0010573	0.00012935	6.5268e-05	0.00013172	3.9161e-05	0.005028	0.00081763	-	8.3069e-06	-	1.3054e-05	0.00016851	-	-	-	0.00029074	5.9335e-06	-	1.1867e-05
0.0010751	4.7468e-06	1.424e-05	2.3734e-06	4.7468e-06	0.0025514	1.1867e-06	2.3734e-06	0.0016246	-	0.00054351	0.00063013	1.424e-05	0.0004818	0.00021598	7.1202e-06	4.7468e-06	5.9335e-06	2.3734e-06	-	-	1.1867e-06	1.1867e-06	1.1867e-06	-	-	-	-	-	-	2.3734e-06	
0.0004367	-	-	-	-	0.0016317	-	-	0.00063488	-	-	-	-	-	5.8148e-05	-	-	4.7468e-06	-	-	0.00017919	-	-	-	-	-	-	-	-	4.6281e-05	-	-
Ы	5.9335e-05	0.00039636	0.0014501	0.00021123	0.00046162	0.0010004	7.7135e-05	0.00022547	0.00047349	0.0014857	0.00039873	0.0023141	0.0012982	0.0007405	0.0003121	0.00053876	0.00036194	0.0012342	0.0012496	0.00015308	4.7468e-06	0.00098377	5.9335e-06	0.0003477	0.00055181	1.068e-05	-	-	0.00010443	1.1867e-06	5.9691e-05
0.00030023	0.00051859	0.00143	0.00032278	0.00064912	0.0010182	0.00010918	0.00052571	0.0014276	-	0.0023461	0.00021479	0.00083662	0.0041285	0.00096716	0.0012555	0.00031803	0.0021942	0.0010395	0.00031329	4.6281e-05	0.00031329	0.00011036	0.00076423	0.00037618	3.7974e-05	-	-	0.00023734	0.00038449	0.0010372	
Э	-	-	7.1202e-06	1.1867e-06	9.4936e-06	-	-	-	1.5427e-05	4.9841e-05	5.9335e-06	3.5601e-06	2.8481e																		

Энтропия: 4.1271

Duration: 3.0433

Лабораторная работа №1

calH(30): $1.9987 < H < 2.6226$

Лабораторная работа №1

Произвольная часть текста:
дц_справедливыми_и_несправедливыми_договорами_я_думаю_шила_в_мешке_не_утаиш

Использованные буквы:

Порядок n-граммы:
5 символов
10 символов
15 символов
20 символов
25 символов
30 символов
35 символов
40 символов
45 символов
50 символов

Введенный символ: и
Символ по счету: 1
Номер эксперимента: 50
Поле ввода символов:
и
Продолжить Другой

Неравенство для энтропии:
1,99867595623896 < H < 2,62262668981542
Двоичная таблица угаданных символов:
10000000000000000000000000000000 ^
00100000000000000000000000000000
0000000000000000000001000000000000
10000000000000000000000000000000
01000000000000000000000000000000 v

Вероятности:
q[1] = 0,46
q[2] = 0,24
q[3] = 0,02
q[4] = 0,02
q[5] = 0
q[6] = 0,02
q[7] = 0
q[8] = 0
q[9] = 0
q[10] = 0,02
q[11] = 0
q[12] = 0,02
q[13] = 0
q[14] = 0,02
q[15] = 0,02
q[16] = 0,02
q[17] = 0,02
q[18] = 0,04
q[19] = 0,04
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0
q[24] = 0
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0,02
q[29] = 0
q[30] = 0,02
q[31] = 0
q[32] = 0

Строка состояния:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

Оцінка надлишковості:

	Текст з пробілами, R	Текст без пробілів, R
H1	0.12218	0.11026
H2	0.21136	0.17472
H2 (з кроком 2)	0.21134	0.17458

CoolPinkProgram:

	R
H10	$0.6433 > R > 0.48982$
H20	$0.71596 > R > 0.58588$
H30	$0.60026 > R > 0.47548$

Висновки:

Для отриманих значень було використано алфавіт російської мови в якому замінені букви «ъ» «ё» на аналогічні. Довжина вхідного алфавіту 32 символу.

Для даної мови (російської) розраховано, що надлишковість встановлює 0.18163.

Встановлена залежність: при збільшенні n питомої ентропії її значення зменшується. Так аналітично можливо визначити наступні члени, оскільки функція є неперервною.