

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

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1
З дисципліни «Криптографія»
Експериментальна оцінка ентропії на символ джерела
відкритого тексту

Виконали:
Правдива Тамара ФБ-02
Бобер Наталія ФБ-05

Перевірила:
Селюх П. В.

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.

Хід роботи

Для виконання цієї лабораторної роботи в якості експериментального тексту було використано декілька глав «Стив Джобс - Уолтер Айзексон».

Перш за все, була створена програма, яка очищує текст від непотрібних символів (цифри, знаки пунктуації, тощо). Лише після цього була створена програма для обробки тексту та виконання завдання. Так як необхідно працювати як з пробілом, так і без нього, робота програми відбувається в два етапи, які відрізняються один від одного лише алфавітом.

Якщо говорити про труднощі, то найбільше їх виникло при роботі з біграмами, де потрібно було враховувати перетин (або його відсутність) букв, який впливав на підрахунок частот та ентропії, вивід результатів в таблиці, де так і не вдалося вивести числа з пробілами в текстовику, напевно, зручніше було б вивести в таблицю excel, але і з цим були певні проблеми.

Результати

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.

Результати з пробілами:

H letter: 4.381850329932886

H cross bigrams: 4.000506233421664

H non-cross bigrams: 4.000363326572691

Результати без пробілів:

H letter: 4.45483115391814

H cross bigrams: 4.137830453960522

H non-cross bigrams: 4.137314968244347

Результати з пробілами

“” :0.157350
о:0.096570
е:0.066070
а:0.065950
и:0.060730
н:0.054390
т:0.051820
с:0.047590
л:0.043950
р:0.039790
в:0.036190
д:0.028340
к:0.028270
м:0.027250
п:0.024990
у:0.021090
б:0.017080
ь:0.015470
я:0.015460
ы:0.014840
з:0.013860
г:0.013570
ч:0.012050
ж:0.010260
й:0.008380
х:0.006470
ш:0.005380
ю:0.005110
ц:0.003620
э:0.003360
ф:0.002370
щ:0.002130
ъ:0.000260
ё:0.000000

Результати без пробілів

о:0.114600
е:0.078400
а:0.078270
и:0.072070
н:0.064540
т:0.061490
с:0.056480
л:0.052150
р:0.047220
в:0.042940
д:0.033630
к:0.033550
м:0.032340
п:0.029650
у:0.025030
б:0.020270
ь:0.018360
я:0.018340
ы:0.017610
з:0.016450
г:0.016110
ч:0.014310
ж:0.012170
й:0.009940
х:0.007670
ш:0.006380
ю:0.006060
ц:0.004300
э:0.003990
ф:0.002820
щ:0.002530
ъ:0.000310
ё:0.000000

Результати з пробілами (cross)

	а	б	в	г	д	е	ё	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я	“ ”	
а	0.00001	0.00137	0.00304	0.00059	0.00142	0.00103	0.00000	0.00123	0.00332	0.00024	0.00073	0.00351	0.01030	0.00265	0.00500	0.00004	0.00088	0.00250	0.00360	0.00510	0.00035	0.00023	0.00062	0.00049	0.00078	0.00053	0.00015	0.00000	0.00000	0.00000	0.00000	0.00005	0.00044	0.00133	0.01438
б	0.00069	0.00006	0.00003	0.00000	0.00001	0.00166	0.00000	0.00000	0.00001	0.00103	0.00000	0.00009	0.00070	0.00005	0.00024	0.00250	0.00000	0.00125	0.00333	0.00001	0.00085	0.00000	0.00008	0.00000	0.00000	0.00028	0.00021	0.00343	0.00001	0.00001	0.00002	0.00033	0.00019		
в	0.00593	0.00001	0.00003	0.00001	0.00019	0.00388	0.00000	0.00000	0.00028	0.00267	0.00000	0.00029	0.00082	0.00028	0.00093	0.00578	0.00029	0.00069	0.00235	0.00032	0.00069	0.00000	0.00002	0.00001	0.00002	0.00051	0.00002	0.00001	0.00213	0.00011	0.00000	0.00000	0.00019	0.00774	
г	0.00103	0.00000	0.00000	0.00002	0.00136	0.00062	0.00000	0.00000	0.00000	0.00081	0.00000	0.00004	0.00106	0.00001	0.00017	0.00645	0.00000	0.00109	0.00001	0.00002	0.00031	0.00000	0.00000	0.00000	0.00002	0.00000	0.00000	0.00000	0.00000	0.00001	0.00000	0.00000	0.00054		
д	0.00451	0.00004	0.00047	0.00001	0.00012	0.00442	0.00000	0.00365	0.00006	0.00280	0.00000	0.00016	0.00109	0.00004	0.00159	0.00339	0.00013	0.00101	0.00068	0.00009	0.00163	0.00000	0.00005	0.00012	0.00009	0.00005	0.00000	0.00001	0.00055	0.00019	0.00005	0.00002	0.00016	0.00114	
е	0.00021	0.00105	0.00108	0.00265	0.00280	0.00109	0.00000	0.00052	0.00102	0.00009	0.00195	0.00152	0.00549	0.00386	0.00766	0.00023	0.00061	0.00701	0.00393	0.00476	0.00007	0.00011	0.00069	0.00030	0.00103	0.00080	0.00054	0.0000	00.00000	0.00000	0.00001	0.00007	0.00012	0.01478	
ё	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
ж	0.00076	0.00004	0.00000	0.00000	0.00077	0.00273	0.00000	0.00000	0.00000	0.00121	0.00000	0.00010	0.00001	0.00000	0.00095	0.00331	0.00000	0.00000	0.00001	0.00000	0.00016	0.00000	0.00000	0.00000	0.00001	0.00000	0.00000	0.00000	0.00000	0.00002	0.00000	0.00000	0.00016		
з	0.00440	0.00012	0.00097	0.00022	0.00083	0.00040	0.00000	0.00006	0.00010	0.00056	0.00000	0.00010	0.00019	0.00042	0.00162	0.00074	0.00001	0.00038	0.00001	0.00000	0.00040	0.00000	0.00000	0.00001	0.00000	0.00000	0.00000	0.00001	0.00061	0.00012	0.00000	0.00000	0.00015	0.00144	
и	0.00045	0.00033	0.00294	0.00064	0.00126	0.00244	0.00000	0.00016	0.00278	0.00129	0.00101	0.00211	0.00641	0.00270	0.00394	0.00055	0.00027	0.00125	0.00332	0.00422	0.00002	0.00017	0.00127	0.00045	0.00122	0.00049	0.00015	0.00000	0.00000	0.00000	0.00002	0.00072	0.00184	0.01632	
й	0.00000	0.00000	0.00002	0.00001	0.00017	0.00002	0.00000	0.00000	0.00004	0.00000	0.00000	0.00015	0.00005	0.00004	0.00051	0.00009	0.00001	0.00001	0.00046	0.00041	0.00000	0.00001	0.00002	0.00001	0.00006	0.00006	0.00000	0.00000	0.00000	0.00000	0.00000	0.00002	0.00621		
к	0.00582	0.00000	0.00014	0.00000	0.00000	0.00079	0.00000	0.00008	0.00001	0.00226	0.00000	0.00022	0.00087	0.00001	0.00028	0.00924	0.00000	0.00133	0.00023	0.00127	0.00168	0.00000	0.00001	0.00028	0.00000	0.00001	0.00000	0.00000	0.00000	0.00000	0.00001	0.00000	0.00363		
л	0.00572	0.00001	0.00006	0.00008	0.00003	0.00406	0.00000	0.00031	0.00001	0.00719	0.00000	0.00030	0.00146	0.00009	0.00029	0.00500	0.00002	0.00000	0.00198	0.00006	0.00114	0.00001	0.00002	0.00000	0.00005	0.00001	0.00000	0.00000	0.00036	0.00414	0.00002	0.00099	0.00177	0.00880	
м	0.00304	0.00004	0.00003	0.00001	0.00002	0.00349	0.00000	0.00000	0.00001	0.00306	0.00000	0.00006	0.00014	0.00034	0.00128	0.00288	0.00142	0.00002	0.00017	0.00003	0.00200	0.00001	0.00000	0.00001	0.00002	0.00000	0.00001	0.00000	0.00017	0.00009	0.00002	0.00004	0.00037	0.00748	
н	0.00958	0.00009	0.00010	0.00021	0.00082	0.00755	0.00000	0.00019	0.00007	0.00836	0.00000	0.00043	0.00001	0.00000	0.00245	0.00908	0.00001	0.00022	0.00081	0.00148	0.00160	0.00020	0.00004	0.00056	0.00020	0.00005	0.00004	0.00000	0.00338	0.00086	0.00000	0.00011	0.00149	0.00440	
о	0.00015	0.00731	0.00732	0.00459	0.00524	0.00184	0.00000	0.00145	0.00175	0.00103	0.00275	0.00178	0.00509	0.00729	0.00687	0.00024	0.00118	0.00554	0.00576	0.00725	0.00023	0.00027	0.00036	0.00023	0.00144	0.00060	0.00023	0.00000	0.00000	0.00000	0.00017	0.00012	0.00063	0.01785	
п	0.00172	0.00002	0.00000	0.00000	0.00000	0.00218	0.00000	0.00000	0.00000	0.00107	0.00000	0.00008	0.00079	0.00000	0.00010	0.00925	0.00016	0.00742	0.00010	0.00006	0.00080	0.00000	0.00000	0.00003	0.00001	0.00000	0.00000	0.00000	0.00006	0.00028	0.00069	0.00001	0.00000	0.00005	
р	0.00870	0.00006	0.00052	0.00029	0.00028	0.00676	0.00000	0.00038	0.00003	0.00496	0.00000	0.00051	0.00002	0.00032	0.00100	0.00784	0.00019	0.00018	0.00047	0.00062	0.00237	0.00008	0.00008	0.00021	0.00007	0.00019	0.00001	0.00000	0.00144	0.00029	0.00008	0.00005	0.00067	0.00112	
с	0.00215	0.00007	0.00105	0.00002	0.00033	0.00296	0.00000	0.00000	0.00000	0.00174	0.00000	0.00382	0.00205	0.00067	0.00089	0.00356	0.00208	0.00029	0.00117	0.01136	0.00081	0.00004	0.00025	0.00011	0.00032	0.00008	0.00000	0.00003	0.00030	0.00251	0.00011	0.00008	0.00395	0.00482	
т	0.00571	0.00003	0.00232	0.00001	0.00014	0.00581	0.00000	0.00000	0.00001	0.00475	0.00000	0.00075	0.00030	0.00012	0.00107	0.01270	0.00014	0.00303	0.00113	0.00028	0.00115	0.00004	0.00000	0.00018	0.00010	0.00001	0.00001	0.00000	0.00114	0.00591	0.00004	0.00005	0.00043	0.00444	
у	0.00009	0.00038	0.00054	0.00085	0.00165	0.00019	0.00000	0.00096	0.00033	0.00005	0.00006	0.00072	0.00096	0.00093	0.00040	0.00004	0.00083	0.00067	0.00132	0.00092	0.00000	0.00001	0.00029	0.00001	0.00112	0.00042	0.00020	0.00000	0.00000	0.00000	0.00015	0.00105	0.00004	0.00591	
ф	0.00011	0.00000	0.00000	0.00000	0.00000	0.00051	0.00000	0.00000	0.00000	0.00063	0.00000	0.00000	0.00006	0.00005	0.00000	0.00047	0.00000	0.00027	0.00002	0.00003	0.00008	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00004		
х	0.00053	0.00000	0.00012	0.00001	0.00001	0.00030	0.00000	0.00000	0.00000	0.00023	0.00000	0.00001	0.00007	0.00004	0.00028	0.00187	0.00000	0.00005	0.00003	0.00001	0.00015	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00001	0.00004	0.00000	0.00000	0.00272	
ц	0.00033	0.00000	0.00009	0.00000	0.00001	0.00099	0.00000	0.00000	0.00002	0.00133	0.00000	0.00002	0.00000	0.00000	0.00000	0.00016	0.00000	0.00000	0.00000	0.00001	0.00011	0.00019	0.00000	0.00001	0.00000	0.00000	0.00000	0.00000	0.00005	0.00000	0.00000	0.00000	0.00028		
ч	0.00164	0.00000	0.00001	0.00000	0.00000	0.00309	0.00000	0.00000	0.00000	0.00185	0.00000	0.00017	0.00005	0.00000	0.00091	0.00003	0.00000	0.00002	0.00000	0.00335	0.00044	0.00000	0.00000	0.00000	0.00000	0.00019	0.00000	0.00000	0.00000	0.00016	0.00000	0.00000	0.00000	0.00016	
ш	0.00046	0.00000	0.00001	0.00000	0.00000	0.00166	0.00000	0.00000	0.00000	0.00133	0.00000	0.00033	0.00047	0.00001	0.00025	0.00017	0.00000	0.00004	0.00000	0.00013	0.00017	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00029	0.00000	0.00000	0.00005		
щ	0.00042	0.00000	0.00000	0.00000	0.00000	0.00096	0.00000	0.00000	0.00000	0.00056	0.00000	0.00000	0.00000	0.00007	0.00000	0.00000	0.00001	0.00000	0.00000	0.00006	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00005	0.00000	0.00000	0.00000		
ъ	0.00000	0.00000	0.00000	0.00000	0.00000	0.00008	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00018	0.00000		
ы	0.00000	0.00015	0.00108	0.00015	0.00011	0.00114	0.00000	0.00002	0.00007	0.00002	0.00164	0.00014	0.00202	0.00120	0.00018	0.00001	0.00023	0.00012	0.00054	0.00049	0.00001	0.00000	0.00096	0.00000	0.00019	0.00033	0.00001	0.00000	0.00000	0.00000	0.00000	0.00004	0.00399		
ь	0.00000	0.00007	0.00005	0.00009	0.00025	0.00023	0.00000	0.00000																											

Результати з пробілами (non-cross)

	а	б	в	г	д	е	ё	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я	" "
а	0.000010.001380.003020.000630.001430.000990.000000.001160.003250.000230.000730.003450.010460.002700.005050.000040.000890.002540.003630.004980.000360.000220.000610.000490.000760.000530.000130.000000.000000.000000.000060.000470.001290.01463																																	
б	0.000730.000050.000030.000000.000010.001630.000000.000000.000020.001080.000000.000070.000680.000050.000240.002530.000000.000840.000000.000080.000000.000000.000000.0000320.000200.000330.000010.000010.000040.0000350.00019																																	
в	0.005780.000010.000030.000000.000010.00180.003840.000000.000000.000270.002580.000000.000320.000890.000290.000920.005690.000280.000770.002250.000310.000680.000000.000020.00010.00010.000550.00010.00010.002140.000100.000000.000000.000220.00763																																	
г	0.001000.000000.000010.000030.001360.000620.000000.000000.000000.000000.000080.000000.000040.01010.000000.000160.006510.000000.001080.000000.000020.000300.000000.000000.000000.000000.000020.000000.000000.000000.000000.000000.000000.000000.000000.000052																																	
д	0.004590.000050.000040.000020.000120.004320.000000.000350.000090.002710.000000.000190.001050.000040.001650.003430.000120.001060.000640.000110.001500.000000.000040.000100.000100.000050.000000.000010.000520.000190.000060.000030.000140.00119																																	
е	0.000230.001070.001090.002600.002810.001140.000000.000550.001020.000090.001970.001530.005600.003800.007790.000250.000590.006960.003990.004980.000090.000120.000710.000310.001070.000770.000530.000000.000000.000000.000010.000070.000130.01445																																	
ё	0.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000																																	
ж	0.000770.000030.000000.000000.000800.002650.000000.000000.000000.000000.001240.000000.000080.000000.000000.000090.003390.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000015																																	
з	0.004310.000130.001010.000240.000860.000410.000000.000050.000110.000530.000000.000100.000220.000430.001610.000730.000000.000360.000010.000000.000420.000000.000000.000010.000010.000000.000000.000000.000640.000130.000000.000000.000160.00156																																	
и	0.000470.000040.002870.000640.001320.002500.000000.000150.002680.001200.001120.002070.006340.002750.003910.000560.000250.01260.003160.004090.000020.00140.001310.000510.001240.000480.00140.000000.000000.000000.000000.000010.000780.001980.01626																																	
й	0.000000.000000.000020.000010.000180.000010.000000.000000.000000.000000.000000.000000.000140.000050.000040.000510.000070.000000.000010.000500.000000.000020.000020.000010.000400.000600.000000.000000.000000.000000.000000.000000.000000.000010.00620																																	
к	0.005780.000000.000130.000000.000000.000820.000000.000090.000000.002190.000000.000250.000840.000300.009160.000000.001290.000250.001230.001730.000000.000000.000260.000000.000010.000000.000000.000000.000000.000130.000000.000000.000366																																	
л	0.005710.000010.000060.000070.000030.004170.000000.000300.000000.000730.000000.000290.001350.000080.000290.004950.000030.000000.001920.000070.001180.000010.000020.000000.000040.000000.000000.000000.000360.004170.000010.001000.001760.00860																																	
м	0.002970.000040.000040.000010.000030.000390.000000.000000.000000.000300.000000.000660.000160.000330.001370.002880.001380.000030.000190.000030.001990.000010.000000.000000.000000.000000.001170.000100.000010.000050.000370.00755																																	
н	0.009570.000090.000110.000020.000810.007400.000000.000220.000090.008280.000000.000450.000020.000010.002430.009060.000220.000570.000890.001390.001590.000200.000040.000570.000250.000060.000040.000000.000330.000910.000000.000100.001480.00423																																	
о	0.000140.007160.007510.004530.005110.001890.000000.001450.00																																	

Результати без пробілів (cross)

	а	б	в	г	д	е	ё	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
а	0.000230.002210.005180.001010.002980.001630.000000.001530.004350.001410.000870.005320.012470.003850.007330.001300.002730.003600.005940.006690.000810.000420.008900.000610.001270.000690.000180.000000.000000.000000.000040.000530.001776	0.000820.000080.005050.000010.000020.001970.000000.000000.000000.000010.001250.000000.000110.000830.000070.000300.002980.000010.001480.003970.000020.001020.000000.000090.000000.000000.000000.000030.000250.004070.000090.000030.00039	0.007200.000300.000540.000440.000770.004750.000000.000070.000480.003680.000000.001250.001130.000680.001750.007600.001160.001110.003720.000880.001010.000060.000080.000100.000280.000690.000020.000010.002520.000130.000170.000010.00035	0.01230.00040.00080.000030.001670.000760.000000.000000.000030.001010.000000.000080.001260.000030.000240.007700.001320.000070.000030.000390.000000.000000.000000.000000.000000.000000.000000.000010.000000.00001	г	0.005360.000090.000670.000060.000250.005260.000000.000430.000110.000340.000000.000300.001320.000090.002060.004090.000300.001270.000940.000160.001990.000010.000070.000140.000120.000060.000000.000010.000650.000230.000070.000030.00020	е	0.000440.001950.002930.003530.004350.001540.000000.000740.001750.001200.002320.002780.006780.005230.010440.001610.002940.008930.006350.006260.000640.000210.001100.000420.001680.001030.000650.000000.000000.000000.000020.000090.00027	0.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.00000	ж	0.000900.000050.000020.000100.000940.003240.000000.000000.000000.000010.001450.000000.000130.000010.001010.001150.003950.000020.000110.000020.000000.000200.000000.000000.000000.000000.000020.000000.000000.000000.000000.000000.00000	з	0.005240.000000.001310.000290.001060.000530.000000.000080.000160.000760.000000.000280.000250.003550.002080.000990.000190.000510.0000210.000070.0000510.0000110.000010.000010.000000.000000.000000.000000.000000.000010.000730.000150.000030.000000.00019	и	0.000770.001000.005480.001130.002800.003270.000000.000310.003760.000700.001200.003620.007820.000830.006320.002110.002750.00180.005830.005670.000530.000320.001750.000570.001930.000630.000180.000000.000000.000000.000000.000860.00238	й	0.000120.000260.000600.000190.00750.000090.000000.000090.000170.000580.000000.000850.000150.000370.001080.000590.000850.000320.001260.00720.000190.000110.000050.000070.000200.000110.000010.000000.000000.000000.000090.000010.00007	к	0.006960.000140.000460.000070.000320.000090.000000.000010.000070.003100.000000.000480.001100.000190.0007010.001120.000450.001760.000000.000710.0002100.000320.000050.000340.000160.000010.000000.000000.000000.000000.000200.000010.00005	л	0.009630.000310.001200.000280.000800.005270.000000.000410.000170.0009030.000000.000930.001850.000440.001140.006830.000890.000230.003340.000340.001540.000050.000120.000020.001140.000020.000000.000000.000430.004920.000170.001180.00216	м	0.003740.000320.000770.000220.000680.004270.000000.000070.000200.004380.000000.000660.000270.000660.002280.004090.002600.000310.001120.000250.002610.000070.000090.000030.000400.000020.000010.000000.001390.000110.000150.000060.00052	н	0.011410.000080.000540.000350.001210.000930.000000.000280.000230.010300.000000.000730.000080.000180.003930.011150.000970.000440.001430.01920.002090.000260.000900.000660.000320.000070.000050.000000.004010.001020.000050.000130.000182	о	0.00320.009730.010770.005750.007490.002730.000000.001900.002610.002340.003270.003200.006340.009430.010050.001970.003650.007240.008900.009500.000900.000140.000630.000320.002440.000770.000280.000000.000000.000000.000620.000150.000102	п	0.002040.000030.000000.000000.000000.002580.000000.000000				

Результати без пробілів (non-cross)

	а	б	в	г	д	е	ё	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
а	0.000200.002220.0005290.00110.003050.001650.000000.001590.004350.001410.000810.0005470.012490.003970.001780.001170.002270.003570.0005960.006740.000830.000360.000840.000650.001190.000670.000200.000000.000000.000000.000420.000580.00181																																
б	0.000800.000080.000050.000000.000020.001940.000000.000000.000000.000020.001210.000000.000100.000760.000050.000300.003000.000000.002020.01540.004000.000010.000990.000000.000000.000000.000000.000300.000240.004050.000020.000070.000020.00043																																
в	0.006950.000270.000570.000470.000760.004800.000000.000060.000460.003670.000000.001220.001190.000750.001780.007610.001160.001100.003820.000840.001050.000040.000100.000090.000250.000620.000030.000010.002470.000090.000160.000020.00032																																
г	0.001300.000030.000050.000020.001730.000810.000000.000000.000020.001120.000000.000080.001280.000040.000270.007890.000070.001270.000070.000030.0000440.000000.000000.000000.000050.000000.000000.000000.000000.000010.000000.00000																																
д	0.005450.00100.000680.000090.000230.005210.000000.000430.00110.003500.000000.000290.001390.000070.002010.003990.000290.001160.000910.000150.001960.000000.000050.000140.000110.000080.000000.000020.000590.000240.000110.000020.00018																																
е	0.000470.002010.002810.003360.004330.001650.000000.000770.001730.001180.002320.002780.006540.005180.010510.001480.002900.008810.006460.006210.000620.000200.001120.000410.001660.000960.000690.000000.000000.000000.000000.000250.000070.00024																																
ё	0.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.000000.																																
ж	0.000940.000040.000020.000010.000900.003160.000000.000010.000010.001450.000000.000120.000010.000010.001090.003970.000570.000010.001000.000000.000180.000000.000000.000000.000020.000000.000000.000000.000000.000000.000000.000000.00000																																
з	0.005140.000200.001330.000320.001140.000600.000000.000080.000170.000730.000000.000340.000250.000610.002040.001030.000220.000520.000200.000080.000540.000010.000010.000020.000020.000000.000000.000010.000700.000140.000050.000000.00021																																
и	0.000790.001010.005490.001110.002740.003160.000000.000300.003550.0002720.001210.003630.007800.000880.006370.002120.002840.002180.005810.005710.000490.000360.001830.000610.001840.000590.000180.000000.000000.000000.000000.000370.000850.00239																																
й	0.000110.000250.000620.000200.000730.001100.000000.000090.000170.000570.000000.000080.000150.000390.001090.005570.000790.000320.001230.0																																

2. За допомогою програми CoolPinkProgram оцінити значення (10) H , (20) H , (30) H .

(10) H

Лабораторная работа №1

Произвольная часть текста:
одрозумевали_под_этим_закон_человеческой_природы_их_идея_состояла_в_том_что

Использованные буквы:
ш,

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: а

Символ по счету: 2

Номер эксперимента: 52

Неравенство для энтропии:
 $2,27841234643932 < H < 2,92022337194317$

Двоичная таблица угаданных символов:
01000000000000000000000000000000
01000000000000000000000000000000
10000000000000000000000000000000
00000000000000000000000000000000
01000000000000000000000000000000

Вероятности:
q[1] = 0,4423076
q[2] = 0,1730769
q[3] = 0
q[4] = 0,0384615
q[5] = 0,0384615
q[6] = 0,0192307
q[7] = 0,0384615
q[8] = 0
q[9] = 0
q[10] = 0
q[11] = 0,019230
q[12] = 0,019230
q[13] = 0
q[14] = 0,019230
q[15] = 0
q[16] = 0
q[17] = 0,038461
q[18] = 0
q[19] = 0
q[20] = 0,038461
q[21] = 0
q[22] = 0
q[23] = 0
q[24] = 0,019230
q[25] = 0
q[26] = 0,019230
q[27] = 0,019230
q[28] = 0,019230
q[29] = 0
q[30] = 0,038461
q[31] = 0
q[32] = 0

Поле ввода символов:
а

Продолжить Другой

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

(20) H

Лабораторная работа №1

Произвольная часть текста:
е_данное_вам_но_если_вы_попытуе_нарушить_обещание_данное_ему_то_не_успеет

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: 52

Неравенство для энтропии:
 $2,30785250365036 < H < 3,03751618762915$

Двоичная таблица угаданных символов:
00000010000000000000000000000000
01000000000000000000000000000000
00100000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:
q[1] = 0,4230769
q[2] = 0,0961538
q[3] = 0,0769230
q[4] = 0,0384615
q[5] = 0,0384615
q[6] = 0,0192307
q[7] = 0,0769230
q[8] = 0
q[9] = 0
q[10] = 0,019230
q[11] = 0,019230
q[12] = 0
q[13] = 0
q[14] = 0,038461
q[15] = 0,057692
q[16] = 0
q[17] = 0,019230
q[18] = 0
q[19] = 0
q[20] = 0,038461
q[21] = 0
q[22] = 0
q[23] = 0,019230
q[24] = 0
q[25] = 0
q[26] = 0,019230
q[27] = 0
q[28] = 0
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0

Поле ввода символов:
и

Продолжить Другой

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

$$(30)H$$
[illegible]

3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Результати з пробілами:

R letter: 0.13869634694845956
R cross bigrams: 0.2136539650010565
R non-cross bigrams: 0.21368205500454784

Результати без пробілів:

R letter: 0.11687488159931758
R cross bigrams: 0.17971705698388785
R non-cross bigrams: 0.17981924680172867

Resultate for CoolPinkProgramm:

$$\begin{aligned} 0,4259961275 &< R(10) < 0,5521515503 \\ 0,4029408602 &< R(20) < 0,5463647449 \\ 0,4518147344 &< R(30) < 0,6091357289 \end{aligned}$$

Висновки:

В результаті виконання даної лабораторної роботи ми отримали такі навички:

- о Підрахунку частоти монограм і біграм у тексті
- о Знаходження ентропії для монограм і біграм
- о Визначення надлишковості мови
- о Робота з CoolPinkProgram

При виконанні першого завдання, ми практично переконалися у тому, що літерами з найбільшими частотами є 'о', 'а' та 'е'. В той час як літерами з найменшою частотою є 'ъ' та 'ё'. В третьому пункті лабораторної роботи ми вираховували надлишковість російської мови, побачили її залежність від ентропії та кількості символів в алфавіті.