

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ім. СІКОРСЬКОГО»
ІНН ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ**

**КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1 З ПРЕДМЕТУ «СИМЕТРИЧНА
КРИПТОГРАФІЯ»
«ЕКСПЕРИМЕНТАЛЬНА ОЦІНКА ЕНТРОПІЇ НА СИМВОЛ ДЖЕРЕЛА
ВІДКРИТОГО ТЕКСТУ»**

Виконали:

ФІ-04 Коваль Марія

ФІ-04 Недашківський Іван

Перевірив:

Чорний О.М.

Мета комп'ютерного практикуму

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.

Постановка задачі та варіант 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 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.
- З методичними рекомендаціями були ознайомлені, усі додаткові вказівки виконали. У якості варіанту обрали книгу «Алексей Плущер-Сарно. Большой словарь мата. Том 1», з метою показати справжнє ВЄЛІЧІЄ російської мови :)

Хід роботи, опис труднощів, що виникали, та шляхів їх розв'язання

Переважає більшість завдань комп'ютерного практикуму не викликали труднощів, а навпаки – зацікавленість матеріалом, що дало можливість досконально розібратись у темі. Певні перепони виникали під час форматування результатів, з метою вивести їх у таблицю, а також під час написання алгоритмів обраховування ентропії для біграм. Усі перепони були подолані.

Далі будуть наведені усі результати роботи.

Спершу прикріплюємо таблиці частот букв і біграм відповідного тексту.

1.Біграми, що перетинаються, з пробілами

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я		
а	б	769	2425	549	1052	1304	833	1951	126	539	3375	3862	1773	3821	29	597	3001	2241	4554	98	272	736	353	1317	453	204	0	0	17	595	1210	12253	
б	1445	15	46	8	6	1491	22	37	838	0	105	1106	23	280	1860	0	1097	283	5	931	0	80	1	2	4	275	1179	255	18	3	265	1345	
в	4591	11	45	19	193	3223	1	142	2015	0	253	637	129	863	3981	85	532	1433	396	479	27	20	7	36	188	10	1875	156	1	1	233	7265	
г	753	3	32	15	431	315	0	0	816	0	13	508	25	118	4869	1	1183	7	4	307	0	3	0	9	3	0	0	3	0	0	0	515	
д	3250	12	405	24	27	3443	38	32	1947	0	204	664	54	1097	2064	87	816	345	73	968	0	40	76	17	52	4	322	375	0	37	236	1018	
е	64	1601	1854	1589	2200	781	530	899	148	1542	1989	3742	3084	6747	354	653	4363	3533	3970	124	32	374	342	753	670	225	0	0	2	160	122	11659	
ж	539	11	1	2	501	2184	6	0	957	0	42	9	3	606	284	0	24	47	1	200	0	2	1	122	0	0	0	38	1	1	0	178	
з	2509	133	406	94	1073	272	18	11	294	0	87	246	407	1512	432	1	217	107	19	270	0	1	1	3	0	0	612	190	2	4	89	1332	
и	454	709	1320	368	932	1960	153	2152	873	1826	1912	2598	2436	3145	607	197	942	2028	3548	70	224	1251	725	1635	520	136	4	0	1	239	2445	11819	
й	2	5	4	2	109	4	0	5	0	3	53	18	41	241	15	5	2	482	97	0	12	11	51	47	61	1	0	0	1	0	1	10392	
к	5048	1	146	4	39	500	109	10	3610	0	12	632	8	164	6399	3	802	1215	943	1340	7	19	89	7	17	0	10	3	4	3	2	3854	
л	2959	71	12	85	90	3375	95	2	4629	0	336	242	22	257	4611	8	0	476	51	949	24	9	0	31	7	31	300	3695	1	823	1530	2617	
м	2430	41	38	6	0	2815	2	2	1652	0	34	99	179	782	2562	76	6	74	5	1525	10	9	9	20	1	0	820	29	1	3	453	6374	
н	7334	716	12	79	342	6359	12	151	6614	0	626	0	1	2239	7241	1	39	764	982	1108	88	5	304	145	20	64	3690	775	17	98	798	2178	
о	36	2813	7663	3501	3105	1345	1320	881	422	2548	1292	4432	3197	3077	260	1485	4585	4628	4216	62	182	381	207	1735	489	122	0	0	144	187	427	11582	
п	1064	113	0	0	1	1501	0	0	1326	0	50	488	0	53	6524	99	4724	123	47	534	3	0	10	13	17	1	217	74	0	7	168	253	
р	6081	138	286	226	230	4578	256	32	3636	1	808	252	493	1597	4444	86	41	216	505	2173	15	111	46	39	372	11	703	505	6	111	684	2715	
с	1154	70	663	35	107	2367	1	2	1600	0	3489	2735	596	654	3851	1518	552	1133	8381	1016	127	85	145	108	77	1	243	1012	2	108	2086	2703	
т	3962	22	2029	10	91	4257	2	2	3039	0	433	96	35	1211	6895	18	2366	1271	102	1271	4	30	39	69	2	9	1196	4944	3	55	258	5290	
у	399	432	219	467	1131	720	915	322	74	3620	800	581	427	442	4	925	701	1158	906	6	103	250	31	598	650	224	6	0	16	866	851	4207	
ф	273	0	0	0	0	323	0	0	370	0	2	114	104	5	769	0	350	5	12	211	15	0	0	0	0	0	0	3	92	1	3	1	143
х	456	2	99	0	2	57	0	35	210	0	1	63	39	113	1137	5	102	23	38	5208	0	2	1	5	2	0	0	2	3	1	8	3120	
ц	456	0	27	0	0	836	0	0	1181	0	66	3	5	1	71	0	0	0	1	85	0	0	1	0	1	0	134	0	0	0	6	302	
ч	1534	0	4	0	0	3462	0	0	1538	0	327	37	2	750	67	0	13	0	1356	310	0	0	0	1	53	0	1	290	0	0	0	786	
ш	433	1	15	0	0	1165	0	0	920	0	717	193	13	176	160	15	4	2	100	181	0	2	1	0	0	0	0	601	1	0	0	100	
щ	206	0	0	0	0	1222	0	0	642	0	0	0	1	12	1	0	4	0	0	108	0	0	0	0	0	0	0	0	18	0	0	29	
ы	0	97	540	91	84	1109	14	59	7	1485	355	515	789	104	4	156	522	366	405	1	0	1323	4	113	174	8	0	0	0	0	19	2894	
ь	0	128	20	49	25	651	0	109	114	0	754	5	265	1542	3	6	4	600	164	1	7	16	30	33	297	1	0	0	0	258	542	7387	
э	0	9	28	1	119	0	3	24	0	13	126	57	76	92	2	48	166	18	1005	0	5	5	0	0	4	0	0	0	0	0	18	162	
ю	3	292	2	14	214	10	9	22	0	3	20	29	18	32	0	5	64	87	364	0	2	22	10	166	39	558	0	0	0	10	0	1872	
я	0	26	201	58	387	276	43	397	8	51	274	177	390	388	0	45	90	365	1065	3	0	228	53	147	15	279	0	0	0	128	37	8765	
	2874	4784	10305	2637	5236	2201	1378	3063	7623	34	6445	3198	5422	10601	6824	11285	4084	13531	5696	2512	1637	5719	639	3360	612	79	3	14	1739	166	1408	0	

2.Біграми, що перетинаються, без пробілів

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я		
а	304	1330	3309	743	1549	1515	949	2173	889	540	3956	4125	2478	4786	657	1624	3314	3398	5129	293	388	1931	441	1594	506	226	0	4	186	613	1365	NaN	
б	1480	39	93	36	28	1507	26	47	881	1	159	1130	87	540	1927	102	1136	428	210	967	29	112	3	11	5	275	1181	258	38	3	286	NaN	
в	4754	462	497	214	513	3344	133	259	2381	1	631	802	411	1229	4318	720	910	2440	805	597	189	200	46	195	232	15	1875	156	219	23	276	NaN	
г	772	26	77	24	455	325	2	19	845	1	32	519	45	163	4905	47	1189	86	22	317	4	12	5	14	4	2	0	6	5	0	10	NaN	
д	3284	39	531	45	68	3462	46	50	1992	1	252	695	136	1156	2104	160	838	473	100	1021	13	104	80	30	54	4	322	375	10	39	243	NaN	
е	230	2053	2903	1832	2778	963	645	1234	963	1542	2510	4039	3529	7632	1141	1861	4681	4778	4424	406	161	799	395	1134	711	231	0	0	147	171	213	NaN	
ж	546	21	15	9	513	2187	10	6	965	0	49	10	10	624	291	17	29	62	14	202	3	8	3	122	0	1	0	38	2	1	2	NaN	
з	2541	150	481	133	1125	282	41	71	328	0	153	284	431	1616	508	93	246	339	77	328	5	44	15	37	89	0	612	190	22	5	96	NaN	
и	739	1162	2352	563	1456	2178	272	2450	1622	1826	2438	2844	2950	4282	1336	1327	1227	3213	4151	284	345	1878	779	1861	591	143	4	0	146	248	2562	NaN	
й	258	310	785	214	501	166	145	310	569	3	579	323	543	1099	453	1005	284	1821	636	183	295	463	97	278	107	2	0	0	122	12	102	NaN	
к	5115	136	441	66	206	573	148	73	3941	2	211	697	160	541	6625	366	890	1721	1100	1411	63	181	104	88	28	2	10	3	54	8	36	NaN	
л	3018	190	194	113	239	3432	111	75	4804	0	464	274	105	522	4793	276	50	727	164	1009	54	96	18	111	20	31	300	3695	28	824	1601	NaN	
м	2585	226	626	195	233	2899	57	123	2091	0	355	338	458	1226	2924	660	212	743	222	1646	109	214	26	338	21	1	821	29	97	30	552	NaN	
н	7434	833	212	124	439	6397	33	193	6747	0	732	53	92	2378	7332	178	98	1048	1069	1148	96	161	315	174	27	69	3690	775	28	111	816	NaN	
о	240	3165	8681	3704	3466	1563	1482	1408	968	2549	1878	4946	3684	4132	940	2362	5001	5673	4737	317	355	896	239	2140	525	123	0	1	334	192	623	NaN	
п	1070	119	21	11	5	1504	4	3	1336	0	61	490	24	65	6532	125	4737	154	53	539	35	4	11	16	17	1	217	74	7	7	168	NaN	
р	6309	177	455	263	465	4598	268	61	3728	1	908	289	595	1740	4524	203	694	516	558	2195	24	248	57	65	393	11	703	505	31	112	701	NaN	
с	1227	293	802	182	222	2415	29	31	1725	0	3724	2841	715	846	3979	1790	608	1339	8507	1071	159	169	173	169	99	1	243	1012	39	115	2096	NaN	
т	4079	234	2427	107	308	4377	43	108	3360	0	683	183	272	1712	7179	632	2502	1847	366	1406	41	231	69	237	24	15	1196	4944	67	68	294	NaN	
у	498	593	535	543	1286	801	970	405	405	3621	1076	731	613	866	216	1304	801	1528	1084	72	138	416	43	714	666	228	6	0	49	869	974	NaN	
ф	302	5	10	0	3	325	1	0	375	0	4	118	110	20	773	8	360	33	15	211	17	2	1	0	1	0	3	92	3	3	1	NaN	
х	501	68	301	65	188	83	38	108	403	12	206	163	149	337	1294	316	188	519	188	5255	47	58	43	93	6	0	0	4	52	3	46	NaN	
ц	461	7	47	14	6	837	0	7	1199	0	90	9	34	32	84	20	4	17	8	91	0	4	2	54	2	1	134	0	5	0	7	NaN	
ч	1543	26	92	148	6	3469	0	7	1553	0	385	42	50	828	125	123	17	25	1365	347	2	18	0	11	56	0	1	290	1	0	1	NaN	
ш	435	6	17	5	3	1169	2	2	924	0	724	195	16	182	164	23	7	9	104	184	7	18	1	0	0	0	0	601	2	0	0	NaN	
щ	206	1	1	0	3	1222	1	0	644	0	1	0	3	12	3	1	5	9	1	112	0	0	0	0	0	0	0	0	18	0	0	0	NaN
ы	57	229	777	160	228	1169	36	138	238	1489	487	562	930	462	161	421	586	664	537	58	13	1429	17	194	193	10	0	0	24	3	46	NaN	
ь	157	460	666	151	299	852	83	291	597	0	1187	131	569	2268	386	590	163	1400	470	121	70	516	61	294	318	8	0	1	88	263	611	NaN	
э	1	10	77	4	120	1	4	24	5	13	149	58	78	97	2	95	168	27	1005	0	6	8	0	6	4	0	0	0	1	0	18	NaN	
ю	43	353	149	50	291	49	45	57	132	11	129	79	102	185	113	179	121	292	432	26	43	55	39	202	53	562	0	0	24	11	40	NaN	
я	126	302	1273	215	725	442	136	609	619	52	787	368	678	1224	535	782	330	1292	1458	234	85	459	93	349	48	281	0	0	150	133	111	NaN	
	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

3.Біграми, що не перетинаються, з пробілами

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я	
а	4	368	1209	277	560	633	405	967	57	262	1701	1937	891	1892	12	312	1485	1083	2303	57	127	365	197	640	237	99	0	0	7	301	621	6090
б	731	8	24	4	3	761	11	19	410	0	59	542	12	137	933	0	545	137	3	460	0	46	0	2	2	132	584	126	13	2	131	688
в	2329	7	21	10	103	1616	0	73	1021	0	115	316	67	435	2022	45	254	755	191	249	15	11	2	19	82	7	933	74	0	0	111	3653
г	382	1	17	7	205	147	0	0	426	0	5	268	15	64	2418	1	590	2	4	139	0	2	0	6	3	0	0	0	0	0	0	249
д	1565	5	199	12	13	1739	17	19	920	0	99	321	26	572	1027	48	424	172	41	483	0	20	35	10	35	2	162	189	0	17	101	480
е	33	789	932	804	1093	405	266	451	75	777	986	1871	1506	3365	193	341	2212	1819	1988	60	11	195	168	384	352	120	0	0	1	83	57	5807
ж	246	6	1	1	249	1118	2	0	469	0	20	5	2	305	155	0	12	20	0	94	0	0	1	61	0	0	0	14	0	1	0	88
з	1251	67	206	43	552	124	7	4	143	0	50	121	210	737	218	1	108	58	10	143	0	0	0	0	0	0	317	91	2	3	51	643
и	225	355	654	191	468	975	83	1088	402	891	957	1314	1229	1567	289	109	479	1000	1810	33	104	627	353	837	280	70	3	0	0	135	1202	5948
й	1	3	2	2	50	4	0	3	0	2	35	9	21	121	7	2	0	241	43	0	4	6	23	22	27	0	0	0	0	0	0	5191
к	2529	0	69	3	17	251	47	5	1789	0	4	333	4	87	3214	0	377	606	466	652	3	10	35	4	7	0	5	3	1	3	0	1939
л	1452	34	5	37	45	1663	47	1	2302	0	177	125	8	126	2320	5	0	235	24	481	15	6	0	15	6	15	140	1908	1	408	776	1269
м	1190	16	17	3	0	1389	1	1	845	0	14	53	87	394	1280	25	3	37	2	765	5	5	4	12	1	0	411	15	0	1	222	3244
н	3751	355	7	42	177	3163	9	77	3285	0	313	0	1	1103	3697	0	19	395	478	555	34	3	148	80	8	30	1863	391	10	39	407	1081
о	18	1427	3816	1734	1569	663	663	437	200	1261	636	2190	1610	1495	120	718	2285	2332	2065	29	91	184	96	855	243	59	0	0	73	92	220	5682
п	531	63	0	0	0	747	0	0	698	0	24	259	0	27	3231	62	2336	56	27	258	2	0	5	6	9	1	113	37	0	3	94	140
р	3077	70	142	111	109	2263	136	21	1843	0	408	124	253	774	2261	39	17	114	263	1133	5	56	28	17	175	9	367	263	3	62	333	1327
с	553	31	316	17	58	1150	0	2	761	0	1767	1372	306	345	1958	750	257	583	4063	480	66	47	77	52	37	1	134	526	2	47	1034	1392
т	2033	12	1033	7	51	2107	1	0	1536	0	231	51	16	594	3519	9	1156	637	45	659	3	14	23	33	1	5	580	2460	2	29	128	2643
у	201	237	105	236	577	360	463	160	32	1858	385	278	224	203	3	436	352	572	468	4	47	133	15	277	306	118	4	0	8	435	413	2095
ф	141	0	0	0	0	167	0	0	195	0	2	58	56	1	420	0	200	3	5	109	8	0	0	0	0	0	0	2	49	1	0	81
х	227	2	48	0	1	24	0	18	88	0	0	31	23	48	546	3	52	12	19	2580	0	1	0	2	1	0	0	0	2	1	5	1549
ц	232	0	15	0	0	435	0	0	586	0	30	1	2	1	43	0	0	0	1	45	0	0	0	0	0	0	0	75	0	0	0	141
ч	745	0	1	0	0	1749	0	0	756	0	174	20	0	381	34	0	8	0	686	176	0	0	0	0	30	0	1	137	0	0	0	414
ш	203	1	6	0	0	579	0	0	481	0	373	100	3	85	82	11	3	0	45	96	0	1	0	0	0	0	0	275	1	0	0	56
щ	118	0	0	0	0	588	0	0	305	0	0	0	0	1	4	1	0	2	0	0	57	0	0	0	0	0	0	7	0	0	0	13
ы	0	43	244	35	38	559	12	31	4	736	172	271	387	89	1	74	255	179	197	0	0	674	2	62	80	3	0	0	0	0	14	1462
ь	0	60	10	27	10	335	0	44	68	0	352	3	136	761	0	3	4	340	84	1	5	12	10	15	146	0	0	0	0	128	260	3684
э	0	2	13	1	68	0	2	13	0	5	65	28	36	48	2	22	91	8	456	0	2	1	0	0	2	0	0	0	0	0	11	84
ю	2	138	0	5	108	5	1	11	0	2	11	13	9	15	0	4	35	48	166	0	2	11	6	84	19	297	0	0	0	9	0	922
я	0	15	91	25	185	126	23	201	6	31	145	97	198	193	0	24	48	173	551	3	0	115	22	69	6	143	0	0	0	60	20	4388
	1446	2385	5108	1348	2665	1117	694	1534	3848	21	3227	1581	2676	5312	3455	5637	1985	6820	2889	1245	748	2906	316	1653	304	36	0	8	894	85	723	0

4.Біграми, що не перетинаються, без пробілів

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я		
а	153	693	1657	365	766	749	463	1147	446	270	1960	2089	1217	2358	321	835	1688	1739	2523	156	193	974	222	757	266	108	0	1	90	293	681	NaN	
б	743	23	47	19	11	785	12	26	441	1	86	571	53	265	948	49	537	197	99	478	12	55	2	5	2	136	583	139	21	3	150	NaN	
в	2435	228	254	110	259	1674	78	129	1161	1	329	396	193	622	2166	352	449	1226	405	273	91	112	22	87	132	8	966	75	112	9	145	NaN	
г	394	13	34	12	212	172	1	8	441	1	12	285	22	79	2443	22	611	38	7	159	1	6	4	7	0	0	0	5	3	0	5	NaN	
д	1609	21	281	24	32	1754	22	26	982	1	122	328	70	555	1091	77	413	233	48	504	5	52	44	17	27	4	150	188	6	20	130	NaN	
е	114	996	1420	934	1407	463	315	614	468	777	1256	2022	1749	3809	594	962	2317	2359	2248	192	83	410	207	569	346	117	0	0	72	84	100	NaN	
ж	290	13	4	5	270	1096	9	5	454	0	23	7	2	324	128	11	20	27	3	105	2	4	1	61	0	0	0	19	1	0	0	NaN	
з	1257	70	266	72	551	134	22	43	155	0	82	141	207	777	253	47	112	164	37	176	3	24	6	19	53	0	322	101	14	1	47	NaN	
и	346	590	1160	269	737	1075	128	1232	814	948	1256	1424	1479	2146	699	694	624	1632	2072	137	176	924	394	939	306	67	1	0	73	119	1271	NaN	
й	123	148	395	106	244	86	70	151	276	2	290	156	263	536	225	485	142	895	306	84	152	212	50	136	50	1	0	0	51	3	50	NaN	
к	2571	71	217	32	91	313	76	39	1952	1	110	348	80	281	3315	176	436	857	512	722	36	93	53	44	14	1	6	2	27	4	22	NaN	
л	1482	92	108	60	120	1675	55	29	2368	0	240	156	49	259	2416	118	23	378	73	519	33	57	7	64	9	19	150	1834	15	406	833	NaN	
м	1342	123	307	99	121	1491	27	62	1040	0	180	173	216	615	1435	335	114	360	103	802	56	117	14	166	12	1	418	17	46	17	297	NaN	
н	3709	430	111	61	224	3214	19	87	3476	0	354	28	53	1208	3648	90	42	542	507	580	56	97	181	88	14	34	1810	406	14	55	406	NaN	
о	126	1569	4234	1855	1753	788	744	704	457	1299	955	2436	1819	2030	466	1170	2521	2845	2351	148	190	482	120	1063	255	59	0	0	170	96	314	NaN	
п	532	60	10	5	2	741	4	0	636	0	32	261	14	29	3276	49	2364	78	24	255	17	1	7	8	4	0	113	36	6	4	83	NaN	
р	3093	87	228	137	235	2317	127	30	1931	0	432	147	300	885	2255	97	349	294	289	1061	8	132	36	27	194	7	371	254	15	61	340	NaN	
с	586	152	403	85	117	1208	15	18	869	0	1862	1417	352	403	1992	912	308	650	4210	545	84	81	84	95	58	0	121	510	18	62	1058	NaN	
т	2079	120	1216	40	160	2207	24	54	1601	0	352	92	147	862	3656	333	1260	916	173	740	24	109	30	120	13	6	592	2510	29	33	167	NaN	
у	232	284	290	257	681	390	502	190	191	1831	542	367	309	422	122	657	384	786	550	34	78	200	22	348	328	112	2	0	29	464	498	NaN	
ф	159	4	5	0	2	146	0	0	174	0	2	61	52	9	374	3	165	22	7	110	7	1	1	0	1	0	2	41	2	1	0	NaN	
х	259	37	155	29	91	43	16	57	197	6	108	77	82	157	662	170	93	257	89	2585	22	30	22	42	5	0	0	3	23	2	26	NaN	
ц	224	3	24	7	3	392	0	4	604	0	46	3	12	16	34	9	2	7	3	49	0	1	1	22	1	0	62	0	2	0	4	NaN	
ч	765	13	42	65	3	1758	0	6	810	0	193	19	24	425	69	56	10	12	684	165	2	5	0	7	30	0	0	141	1	0	1	NaN	
ш	221	4	8	2	2	584	0	1	434	0	355	111	8	87	89	9	4	7	45	95	3	8	0	0	0	0	0	309	1	0	0	NaN	
щ	96	1	0	0	0	641	1	0	324	0	0	0	1	5	3	0	2	3	1	63	0	0	0	0	0	0	0	11	0	0	0	NaN	
ы	27	113	389	66	98	584	16	59	128	801	240	275	464	220	92	201	281	334	288	27	9	695	9	105	91	4	0	0	10	2	21	NaN	
ь	85	240	341	88	142	394	35	128	298	0	574	61	290	1142	190	288	83	696	211	57	40	249	35	149	144	5	0	0	44	134	326	NaN	
э	0	3	41	0	73	0	3	10	2	8	72	27	42	40	0	44	81	13	524	0	5	3	0	3	1	0	0	0	0	0	11	NaN	
ю	20	175	72	27	135	19	24	29	60	6	60	42	58	90	60	101	57	145	223	15	20	32	21	104	31	266	0	0	11	4	14	NaN	
я	63	150	629	105	349	209	68	298	307	24	373	171	324	602	283	407	166	634	731	113	37	223	46	173	26	136	0	0	69	69	56	NaN	
	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

5. Частота літер у тексті (без пробілу результат не змінюється)

пробіл	125109	'к'	25000	'ч'	10531
'о'	66324	'у'	22051	'з'	10342
'е'	54106	'м'	20057	'г'	9933
'а'	50315	'д'	17727	'ж'	5760
'п'	47229	'п'	17410	'ш'	4800
'н'	42802	'я'	13897	'ю'	3867
'т'	39011	'ь'	13071	'ц'	3176
'с'	36621	'б'	13025	'ф'	2796
'р'	31397	'й'	11665	'щ'	2243
'в'	28847	'ы'	11318	'э'	1981
'л'	27338	'х'	10734		

Одержані значення H_1 та H_2 , оцінки $H^{(10)}$, $H^{(20)}$, $H^{(30)}$ наведені нижче.

$$H_1 = 4.418118868286257$$

$$H_1 = 4.5052822751655155 \text{ (без пробілів)}$$

З пробілами:

$$H_2 = 4.027991959769378 \text{ (з перетинаннями)}$$

$$H_2 = 4.027518991638357 \text{ (без перетинань)}$$

Без пробілів:

$$H_2 = 4.201017242694961 \text{ (з перетинаннями)}$$

$$H_2 = 4.200354905541525 \text{ (без перетинань)}$$

$$1,54505591577676 < H^{(10)} < 2,35873087898791$$

$$1,55851527970516 < H^{(20)} < 2,24980642383963$$

$$1,4720997217181 < H^{(30)} < 2,24014416732061$$

×

×

```
q[32]=0
```


Лабораторная работа №1

Произвольная часть текста:
то_какая_же_может_быть_разница_между_справедливыми_и_несправедливыми_догово

Использованные буквы:
ф, ы, и,

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

Введенный символ: а

Символ по счету: 4

Номер эксперимента: 50

Неравенство для энтропии:
 $1,4720997217181 < H < 2,24014416732061$

Двоичная таблица угаданных символов:

10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Поле ввода символов:
а

Продолжить Другой

Вероятности:

q[1] = 0.52
q[2] = 0.2
q[3] = 0.06
q[4] = 0.02
q[5] = 0.08
q[6] = 0.02
q[7] = 0.04
q[8] = 0.02
q[9] = 0
q[10] = 0
q[11] = 0
q[12] = 0
q[13] = 0
q[14] = 0
q[15] = 0
q[16] = 0.02
q[17] = 0
q[18] = 0
q[19] = 0
q[20] = 0.02
q[21] = 0
q[22] = 0
q[23] = 0
q[24] = 0
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0

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

Нижче наведена оцінка надлишковості російської мови у різних моделях відкритого тексту.

$0.5283 < R < 0.691$ (оцінка для $H^{(10)}$)

$0.55 < R < 0.6883$ (оцінка для $H^{(20)}$)

$0.552 < R < 0.7056$ (оцінка для $H^{(30)}$)

Висновки

Під час виконання практикуму ми засвоїли поняття ентропії та надлишковості джерела, порівняли різні моделі джерела відкритого тексту для наближеного визначення ентропії. Це ми зробили за допомогою освоєння теоретичного матеріалу та застосування практичних навичок програмування та математичних вмінь. Ми отримали необхідні результати на прикладі книги «Алексей Плущер-Сарно. Большой словарь мата. Том 1», змогли вирахувати необхідні значення ентропії для літер та біграм у тексті в різних варіаціях. Переконались в тому, що найчастіші літери російської мови – ‘о’, ‘е’. Окрім того, ми ознайомились з програмою CoolPinkProgram та за її допомогою отримали оцінки умовної ентропії джерела, що дали нам можливість наближено дізнатись надлишковість російської мови на різних порядках n-грамм.