

Звіт до лабораторної роботи 1 з Симетричної Криптографії

ФІ-03 Буржимський Ростислав, Неждодій Максим

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.

Варіантів немає у цій роботі.

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 і основні проблеми

Для підрахунку символів в текстах використовували мову програмування C++. Через погану підтримку кирилиці у цій мові програмування, довелося застосовувати тип даних *wchar_t*, та *wstring*, які були нам не знайомі і не зручні у використанні. Також різниця коду символу 'я'-'а' давала результат 31, що унеможливлювало використання звичайних масивів. Тому довелося застосовувати *map*(словник), що в подальшому дозволило полегшити пошук кількості біграм.

Отримані таблиці з результатами

Частота літер у вихідному тексті.

а	б	в	г	д	е	є	ж	з	и	й	к	л
65994	13305	34641	14280	24714	68034	0	9666	12570	48030	8388	28248	37779
м	н	о	п	р	с	т	у	ф	х	ц	ч	ш
25308	50832	89025	22092	33552	41520	50124	22077	2946	5301	2505	12027	6054
щ	ъ	ы	ь	э	ю	я						
2466	171	13845	16455	2811	6195	17295						

Частоти біграм з перетином літер

	а	б	в	г	д	е	є	ж	з	и	й	к	л	м	н
а	0	603	2391	837	2586	1761	0	1131	2751	123	645	4146	7044	3141	4380
б	1149	147	36	3	15	1881	0	3	0	561	0	72	876	51	330
в	5886	0	18	3	291	4293	0	0	384	2802	0	186	765	186	714
г	1494	0	3	0	915	198	0	0	3	429	0	237	933	3	159
д	3822	30	834	21	36	4482	0	78	309	2820	0	177	630	27	1581
е	102	1191	1002	3723	3021	1296	0	582	1260	90	1791	1032	4650	4209	6444
є	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ж	1332	6	0	6	738	3699	0	12	0	1380	0	174	9	3	861
з	4464	111	804	288	609	561	0	51	12	339	0	138	210	306	1512
и	150	450	1926	690	1800	1809	0	405	1911	504	942	2166	3924	3024	2721
й	0	0	0	0	144	0	0	0	9	0	0	57	21	42	402
к	7959	0	141	0	0	543	0	45	33	2163	0	45	699	45	783
л	5709	36	9	90	18	4113	0	549	6	5127	0	318	141	3	285
м	2655	30	6	18	0	3717	0	0	0	2022	0	66	78	63	1650
н	9030	12	42	93	417	8712	0	48	24	6858	0	399	3	0	2325
о	0	3498	5523	3990	4566	1923	0	2094	1056	747	3360	1884	5334	4788	6492
п	1314	0	0	0	0	1926	0	0	0	900	0	75	783	0	219
р	6744	57	414	210	432	5469	0	300	39	4632	0	264	42	330	786
с	1503	42	1791	18	264	3045	0	36	6	1083	0	3135	2670	594	795
т	5301	9	1941	0	75	5352	0	0	6	4143	0	618	144	0	1197
у	624	465	858	876	1431	288	0	1299	354	42	78	894	1443	963	150
ф	330	0	0	0	0	138	0	0	0	390	0	867	549	3	0
х	318	0	216	6	0	42	0	0	0	150	0	0	81	36	102
ц	519	0	69	3	0	687	0	0	0	444	0	24	0	0	0
ч	1824	0	15	0	0	3261	0	0	0	1545	0	351	27	0	597
ш	756	0	6	0	0	2187	0	0	0	1248	0	354	339	15	237
щ	417	0	0	0	0	1188	0	0	0	627	0	0	0	0	57
ъ	0	0	0	0	0	39	0	0	0	0	0	0	0	0	0
ы	0	147	666	57	159	765	0	21	72	6	1509	123	1521	1179	252
ь	0	18	39	51	72	348	0	0	183	33	0	1080	0	222	1071
э	0	27	3	0	0	0	0	0	6	0	9	27	45	3	39
ю	3	330	9	9	210	0	0	9	3	3	0	36	942	21	51
я	0	15	309	96	426	213	0	57	183	3	51	234	684	306	576

	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э
а	3	1068	3372	3684	4779	114	1533	795	144	762	834	300	0	0	0	0
б	1854	0	987	108	0	1035	0	39	6	12	9	153	108	3096	12	0
в	5967	198	720	2091	336	450	0	24	6	66	708	18	0	2919	255	0
г	7536	0	1029	18	24	432	0	0	0	30	6	0	0	0	6	0
д	3729	168	1185	258	72	1458	0	21	231	48	42	0	15	366	453	3
е	195	837	5655	4221	5679	150	72	366	285	939	687	621	0	0	0	0
ё	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ж	54	0	0	15	3	195	0	0	0	15	0	0	0	0	24	0
з	465	0	234	63	3	270	0	0	3	12	3	0	24	297	75	0
и	1074	216	546	2532	4065	15	36	1197	648	906	564	117	0	0	0	3
й	9	0	0	300	444	0	3	0	42	135	72	0	0	0	0	0
к	7887	0	1632	57	396	1455	0	3	27	0	3	0	0	0	21	0
л	4569	27	0	1356	33	1257	18	0	0	135	3	0	0	453	5085	0
м	3258	123	36	81	0	1959	6	3	3	24	3	12	0	684	36	21
н	8481	0	39	378	981	2409	30	3	309	147	0	69	0	2727	993	0
о	177	1029	4935	5718	6528	54	66	519	99	1884	747	273	0	0	0	36
п	8931	51	6267	36	141	693	0	0	3	15	0	0	0	318	57	0
р	6498	72	24	132	579	2151	0	72	45	84	162	15	0	1452	690	24
с	2412	1614	159	1035	9627	1077	3	180	36	339	69	0	6	264	2319	0
т	12759	60	2214	1050	105	1386	21	15	96	225	6	21	15	1164	5820	0
у	18	735	363	1047	1326	0	42	339	6	849	414	276	0	0	0	3
ф	123	0	255	0	15	141	3	0	0	3	0	0	0	0	3	27
х	1701	0	117	15	6	84	0	0	0	0	0	0	3	0	0	3
ц	144	0	0	3	0	216	0	0	21	0	0	0	0	105	0	0
ч	81	0	51	0	3270	543	0	3	0	3	150	0	0	0	228	0
ш	198	15	0	0	21	258	3	0	0	0	0	0	0	0	336	0
щ	3	0	0	0	0	108	0	0	0	0	0	0	0	0	42	0
ъ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ы	0	144	192	471	729	3	0	627	6	105	333	30	0	0	0	0
ь	15	12	0	984	165	267	3	0	63	51	468	6	0	0	0	0
э	0	12	6	3	2604	0	0	0	0	0	0	0	0	0	0	0
ю	0	0	99	1086	282	0	0	9	6	96	0	294	0	0	0	3
я	0	126	60	459	1233	0	0	132	24	165	9	207	0	0	0	0
	а	б	в	г	д	е	ё	ж	з	и	й	к	л			
ю	3	330	9	9	210	0	0	9	3	3	0	36	942			
я	0	15	309	96	426	213	0	57	183	3	51	234	684			
	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш			
ю	21	51	0	0	99	1086	282	0	0	9	6	96	294			
я	306	576	0	126	60	459	1233	0	0	132	24	165	9			
	щ	ъ	ы	ь	э	ю	я									
ю	0	0	0	0	3	30	0									
я	207	0	0	0	0	114	87									

Частоти біграм без перетину літер

	а	б	в	г	д	е	є	ж	з	и	й	к	л	м	н
а	306	1056	3798	1125	3156	2205	0	1386	3087	735	645	5043	7311	3681	5826
б	1164	150	45	6	21	1893	0	3	15	567	0	93	882	54	339
в	5982	147	279	201	552	4398	0	39	462	2934	0	651	882	315	1083
г	1497	42	63	15	954	216	0	18	15	447	0	252	951	90	297
д	3831	66	891	48	78	4497	0	93	339	2865	0	273	672	63	1680
е	285	2088	2565	4095	3705	1644	0	1173	1773	855	1791	1842	5043	5187	7890
є	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ж	1335	15	6	12	741	3699	0	15	3	1392	0	183	12	9	906
з	4479	141	975	324	654	582	0	75	57	378	0	285	249	372	1638
и	321	1104	3347	966	2295	2130	0	600	2259	1107	945	2916	4134	3375	3783
й	108	231	585	156	582	96	0	108	159	372	0	549	162	363	831
к	8037	213	600	90	156	660	0	135	120	2379	0	471	780	210	1221
л	5886	264	651	225	186	4605	0	699	141	5367	0	696	327	198	840
м	2847	297	780	225	384	3840	0	171	177	2541	0	558	240	300	2355
н	9078	255	513	177	666	8781	0	84	171	7056	0	507	60	111	2766
о	258	4596	7863	4347	5373	2418	0	2403	1599	1563	3360	2886	5736	5550	8415
п	1314	0	3	0	6	1926	0	0	3	900	0	78	786	3	222
р	6768	93	525	237	498	5478	0	306	60	4686	0	324	57	345	885
с	1575	174	2139	111	537	3129	0	228	126	1173	0	3387	2736	726	1200
т	5421	237	2574	123	297	5586	0	93	177	4386	0	945	279	270	1785
у	762	618	1473	978	1620	369	0	1434	510	459	78	1254	1566	1203	876
ф	330	0	12	0	18	138	0	21	3	390	0	873	549	3	3
х	357	90	384	54	120	102	0	36	66	267	0	192	159	111	303
ц	540	12	84	12	3	693	0	3	3	450	0	51	6	15	27
ч	1824	0	27	0	3	3261	0	0	0	1545	0	351	30	0	603
ш	759	3	15	6	9	2187	0	3	0	1254	0	354	339	15	237
щ	423	0	3	0	0	1188	0	0	0	630	0	6	0	0	60
ъ	0	0	0	0	0	39	0	0	0	0	0	0	0	0	0
ы	81	273	1167	159	390	903	0	69	195	261	1509	300	1608	1356	750
ь	120	282	1212	198	474	807	0	117	429	426	0	1650	240	576	2031
э	0	27	6	0	0	0	0	0	6	3	9	39	45	3	42
ю	66	405	303	87	351	45	0	51	75	129	0	315	990	123	216
я	240	426	1752	303	885	519	0	303	540	513	51	924	948	681	1722

	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь		
а	1053	2733	3654	4905	5460	486	1647	852	186	1179	903	300	0	0	0		
б	1881	3	990	114	6	1050	0	39	6	15	9	153	108	3096	12		
в	6261	750	999	2700	741	534	15	51	21	204	759	18	0	2919	255		
г	7602	78	1047	90	45	471	0	3	3	42	12	0	0	0	6		
д	3801	522	1230	384	99	1485	6	27	234	105	45	0	15	366	453		
е	1449	2676	6066	6006	6426	576	156	519	315	1539	768	630	0	0	0		
ё	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
ж	66	21	0	27	9	201	0	0	0	15	0	0	0	0	24		
з	540	171	291	240	72	300	0	9	12	45	15	0	24	297	75		
и	1995	1584	816	3612	4605	312	117	1272	675	1203	675	126	0	0	0		
й	453	672	255	981	687	162	144	21	69	357	141	3	0	0	0		
к	8172	240	1728	402	618	1536	21	33	45	111	24	0	0	0	21		
л	5127	630	150	1908	270	1374	138	15	9	384	33	6	0	453	5085		
м	3930	1008	177	912	348	2187	51	66	15	327	42	21	0	684	36		
н	8829	576	162	915	1107	2589	51	66	315	228	21	69	0	2727	993		
о	1713	3099	5346	7599	7395	615	237	645	171	2802	810	282	0	0	0		
п	8940	51	6267	45	144	693	0	0	3	15	0	0	0	318	57		
р	6573	150	33	219	609	2181	3	75	45	117	165	15	0	1452	690		
с	2610	1962	234	1275	9891	1194	15	195	39	399	78	0	6	264	2319		
т	13146	630	2316	1572	333	1587	60	39	120	423	27	24	15	1164	5820		
у	411	1212	462	1608	1554	135	75	372	15	1068	426	279	0	0	0		
ф	126	12	255	12	18	141	3	0	0	6	0	0	0	0	3		
х	1869	294	183	288	120	132	15	0	12	51	15	0	3	0	0		
ц	174	18	9	27	6	219	3	0	24	12	0	0	0	105	0		
ч	90	3	51	6	3273	543	33	3	0	3	150	0	0	0	228		
ш	198	24	3	12	21	258	6	0	0	6	0	0	0	0	336		
щ	3	3	0	0	0	108	0	0	0	0	0	0	0	0	42		
ъ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
ы	327	705	282	912	945	147	12	690	6	210	348	30	0	0	0		
ь	663	783	156	1950	594	456	48	39	90	447	501	6	0	0	0		
э	3	15	6	3	2604	0	0	0	0	0	0	0	0	0	0		
ю	210	234	129	1317	378	42	21	12	15	228	30	297	0	0	0		
я	810	1233	255	1479	1746	363	69	258	60	486	57	207	0	0	0		
	а	б	в	г	д	е	ё	ж	з	и	й	к	л	м	н	о	п
э	0	27	6	0	0	0	0	0	6	3	9	39	45	3	42	3	15
ю	66	405	303	87	351	45	0	51	75	129	0	315	990	123	216	210	234
я	240	426	1752	303	885	519	0	303	540	513	51	924	948	681	1722	810	1233
	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я	
э	6	3	2604	0	0	0	0	0	0	0	0	0	0	0	0	0	
ю	129	1317	378	42	21	12	15	228	30	297	0	0	0	30	30	66	
я	255	1479	1746	363	69	258	60	486	57	207	0	0	0	120	120	225	

$$H1 = 4,46704$$

Біграми з перетином літер:

1

$$H_2 = 3,96458$$

Біграми без перетину літер:

$$H_2 = 4,14783$$

Лабораторная работа №1

Произвольная часть текста:
ьше_свободы_выбора_чем_ц_камня_упасть_на_землю_или_не_упасть_будучи_организмиз

Использованные буквы:

Порядок n-граммы:	Введенный символ: д	Неравенство для энтропии:
5 символов 10 символов 15 символов 20 символов 25 символов 30 символов 35 символов 40 символов 45 символов 50 символов	Символ по счету: 1	1,27027041306921 < H < 1,96705795760852
	Номер эксперимента: 51	Двоичная таблица угаданных символов:
	Поле ввода символов: д	01000000000000000000000000000000 ^ 10000000000000000000000000000000 10000000000000000000000000000000 10000000000000000000000000000000 10000000000000000000000000000000 v ~~~~~~
		Вероятности:
Строка состояния:		q[1] = 0,66666666 q[2] = 0,0784313 q[3] = 0,0196078 q[4] = 0,0392156 q[5] = 0,0392156 q[6] = 0 q[7] = 0 q[8] = 0,0196078 q[9] = 0 q[10] = 0,019607 q[11] = 0,019607 q[12] = 0 q[13] = 0,039215 q[14] = 0 q[15] = 0,019607 q[16] = 0 q[17] = 0,039215 q[18] = 0 q[19] = 0 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 q[29] = 0 q[30] = 0 q[31] = 0 q[32] = 0
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка		

Лабораторная работа №1



Произвольная часть текста:
кое_он_не_делает_вам_ничего_плохого_почему_я_должен_уступать_тебе_дай_мне_к

Использованные буквы:
в, с,

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: м
Символ по счету: 3
Номер эксперимента: 51
Поле ввода символов:
м
Продолжить Другой

Неравенство для энтропии:
 $1,26053173925093 < H < 2,05939274380502$
Двоичная таблица угаданных символов:
00010000000000000000000000000000
00000000000000000000000000000001
10000000000000000000000000000000
10000000000000000000000000000000
00000010000000000000000000000000
.....

Вероятности:
q[1] = 0,6274509
q[2] = 0,1176470
q[3] = 0,0392156
q[4] = 0,0588235
q[5] = 0,0392156
q[6] = 0,0196078
q[7] = 0,0196078
q[8] = 0
q[9] = 0,0196078
q[10] = 0
q[11] = 0,019607
q[12] = 0
q[13] = 0
q[14] = 0
q[15] = 0
q[16] = 0
q[17] = 0
q[18] = 0
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0
q[24] = 0,019607
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0,019607

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

[illegible]

Оцінка надлишковості

$$\frac{1.2703+1.2605+1.3291}{3} \leq H_{\infty} \leq \frac{1.9671+2.0593+2.1301}{3}$$

$$1.2866 \leq H_\infty \leq 2.0522 - \text{приблизна оцінка } H_\infty$$

Для підрахунку надлишковості оберемо середнє значення.

$$R = 1 - \frac{H_\infty}{H_0} = 1 - \frac{1.6694}{5.0444} = 0.6691$$

Висновок

У цій лабораторній роботі ми пересвідчилися, що наші навички у програмуванні мовою C++ можна розвивати надалі. Ми оновили свої вміння тестувати програмне забезпечення. Під час тестування, було виявлено, що CoolPinkProgram має назву «Лабораторная работа №1», що містить описку у слові «Лабораторная». Також, перевіривши надлишковість російської мови, ми прийшли до висновку, що трошки більше $\frac{2}{3}$ російської мови є зайвою інформацією, що збігається з нашим уявленням. Успішне порівняння кількості символів у тексті точно показало нам, що букви 'ё' немає у нашому тексті. Отримали практичний доказ того, що зі збільшенням кількості символів у тексті ентропія буде знижуватись.