



Міністерство освіти і науки України
Національний технічний університет України
“Київський політехнічний інститут імені Ігоря Сікорського”
Фізико-технічний інститут

Криптографія

Лабораторна №1

Виконали:
Студенти групи ФБ-82
Козачок Вячеслав
Ілля Кузнєцов
Перевірив:

Київ - 2020

1 Порядок виконання роботи

1. Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.
2. 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.
3. За допомогою програми CoolPinkProgram оцінити значення $H(10)$, $H(20)$, $H(30)$
4. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

2 Методичні вказівки

Звичайні текстові файли містять багато символів окрім власне літер; для обчислення значень ентропій вони повинні пройти попередню фільтрацію: всі символи, окрім текстових, повинні вилучатись або замінюватись на пробіли; прописні літери - замінюватись на відповідні стрічні; послідовність пробілів (або інших розділових знаків, наприклад, символів кінця рядку) повинна трактуватись як один пробіл або вилучатись, якщо пробіл не входить до алфавіту.

При підрахунку частот біграм треба розглядати як пари букв, що перетинаються, так і пари букв, що не перетинаються (тобто рухатися вздовж тексту з кроком 2). Одержані результати не повинні суттєво відрізнятись, однак в першому випадку використовується більше статистики, а тому чисельні дані більш точні. Таблицю частот символів потрібно подавати відсортованою за спаданням частот. Таблицю частот біграм зручно подавати у вигляді квадратної матриці, індексованої першою та другою літерами біграм.

Програма CoolPinkProgram використовує текст, що лежить у допоміжному файлі text. Цей текст написаний російською мовою без знаків пунктуації та великих літер; буква «ё» замінена буквою «е», а «ъ» – буквою «ь». Пробіл також вважається буквою. Таким чином, кількість букв алфавіту $m=32$. При підрахунку $H(10)$, $H(20)$, $H(30)$ виконати не менш ніж 50 експериментів.

3 Частота

3.1 Монограми

3.1.1 Монограми без пробіла

	count	percentage
о:	162385,	11.50640%
е:	123629,	8.76020%
а:	117081,	8.29622%
н:	98121,	6.95273%
и:	93860,	6.65080%
т:	84627,	5.99656%
с:	75105,	5.32185%
л:	70907,	5.02438%
в:	66549,	4.71558%
р:	56282,	3.98807%
к:	48456,	3.43353%
д:	41627,	2.94964%
м:	40517,	2.87098%
у:	38126,	2.70156%
п:	34091,	2.41565%
я:	30439,	2.15687%
ь:	27849,	1.97335%
ы:	26210,	1.85721%
г:	25689,	1.82029%
б:	24717,	1.75142%
ч:	23851,	1.69005%
з:	23117,	1.63804%
ж:	16018,	1.13502%
й:	14860,	1.05296%
ш:	12067,	0.85505%
х:	10981,	0.77810%
ю:	8811,	0.62434%
э:	5018,	0.35557%
щ:	4054,	0.28726%
ц:	3992,	0.28287%
ф:	1779,	0.12606%
ъ:	412,	0.02919%
ё:	31,	0.00220%

Entropy:	4.44178288163354	
Redundancy	0.8654005187383775	

3.1.2 Монограми з пробілом

	count	percentage
:	281670,	16.63804%
о:	162385,	9.59196%
е:	123629,	7.30267%
а:	117081,	6.91589%
н:	98121,	5.79593%
и:	93860,	5.54424%
т:	84627,	4.99885%
с:	75105,	4.43640%
л:	70907,	4.18842%
в:	66549,	3.93100%
р:	56282,	3.32454%
к:	48456,	2.86226%
д:	41627,	2.45888%
м:	40517,	2.39331%
у:	38126,	2.25207%
п:	34091,	2.01373%
я:	30439,	1.79801%
ь:	27849,	1.64502%
ы:	26210,	1.54821%
г:	25689,	1.51743%
б:	24717,	1.46001%
ч:	23851,	1.40886%
з:	23117,	1.36550%
ж:	16018,	0.94617%
й:	14860,	0.87777%
ш:	12067,	0.71279%
х:	10981,	0.64864%
ю:	8811,	0.52046%
э:	5018,	0.29641%
щ:	4054,	0.23947%
ц:	3992,	0.23580%
ф:	1779,	0.10508%
ъ:	412,	0.02434%
ё:	31,	0.00183%

Entropy:	4.3521146215748905,	
Redundancy:	0.8719966287772091	

3.2.3 Біграми з пробілами та без перетинів

Bigrams with spaces not intersected ---																																																																	
	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	ъ	ы	ь	э	ю	я	ё																																
а	0.0000	0.3203	0.6599	1.4611	0.3389	0.7130	0.5161	0.2501	0.4163	1.0927	0.0002	0.8964	0.3543	0.5527	1.6213	1.3510	1.4083	0.3842	1.6098	0.7838	0.4102	0.0504	0.1513	0.0262	0.6186	0.0680	0.0099	0.0000	0.0000	0.0000	0.2848	0.0033	0.2203	0.0001																															
б	1.8558	0.0101	0.0581	0.3046	0.0560	0.2180	0.1041	0.1081	0.4109	0.0137	0.0540	0.4143	0.9533	0.2358	0.4637	0.0007	0.0625	0.2301	0.3411	0.4167	0.0544	0.0236	0.0781	0.0123	0.0789	0.0598	0.0191	0.0000	0.0000	0.0000	0.0068	0.0859	0.2398	0.0000																															
в	0.0395	0.0489	0.0001	0.0087	0.0004	0.0037	0.1958	0.0006	0.0065	0.0751	0.0000	0.0277	0.0815	0.0065	0.0286	0.2006	0.0000	0.1067	0.0079	0.0000	0.0958	0.0000	0.0057	0.0004	0.0007	0.0111	0.0249	0.0111	0.0417	0.0017	0.0000	0.0067	0.0487	0.0000																															
г	0.5462	0.5491	0.0015	0.0020	0.0022	0.0295	0.4228	0.0000	0.0538	0.4475	0.0000	0.0157	0.0763	0.0165	0.1259	0.7019	0.0194	0.0147	0.1017	0.3395	0.0236	0.0651	0.0000	0.0058	0.0030	0.0078	0.0015	0.0008	0.0000	0.2425	0.0010	0.0000	0.0010	0.0249	0.0000																														
д	0.0672	0.0816	0.0001	0.0007	0.0004	0.1139	0.0365	0.0005	0.0001	0.0740	0.0000	0.0122	0.1485	0.0000	0.0188	0.8267	0.0000	0.0490	0.0032	0.0010	0.0590	0.0001	0.0000	0.0000	0.0034	0.0070	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000																															
е	0.9420	0.0106	0.0024	0.0737	0.0005	0.0004	0.4390	0.0014	0.0000	0.2072	0.0000	0.0203	0.0795	0.0070	0.1500	0.3659	0.0070	0.0496	0.0318	0.0163	0.1760	0.0000	0.0056	0.0171	0.0268	0.0105	0.0000	0.0066	0.0520	0.0732	0.0000	0.0007	0.0380	0.0000																															
ж	1.8909	0.0060	0.1284	0.2480	0.3654	0.2466	0.2105	0.0769	0.1138	0.0091	0.2477	0.1718	0.5677	0.0404	0.7143	0.0183	0.1377	0.5179	0.3730	0.4535	0.0091	0.0100	0.0008	0.0736	0.0157	0.0878	0.0912	0.0611	0.0000	0.0000	0.0000	0.0400	0.0338	0.0000																															
з	0.0304	0.0407	0.0045	0.0000	0.0006	0.0774	0.4130	0.0111	0.0000	0.1458	0.0000	0.0083	0.0064	0.0000	0.0793	0.0046	0.0000	0.0000	0.0000	0.0000	0.0240	0.0000	0.0000	0.0010	0.0044	0.0000	0.0000	0.0000	0.0000	0.0037	0.0000	0.0010	0.0000	0.0000																															
и	0.0965	0.5290	0.0158	0.0823	0.0541	0.1737	0.0211	0.0093	0.064	0.0289	0.0000	0.0082	0.0197	0.0273	0.1898	0.0372	0.0139	0.0195	0.0565	0.0010	0.0171	0.0235	0.0000	0.0000	0.0000	0.0063	0.0000	0.0000	0.0000	0.0013	0.0370	0.0161	0.0000	0.0067	0.0442	0.0000																													
й	1.8033	0.0090	0.0040	0.2614	0.0409	0.0576	0.2316	0.0313	0.1819	0.0056	0.1336	0.1565	0.4344	0.2776	0.4051	0.0096	0.0198	0.0542	0.2139	0.3762	0.0010	0.0111	0.0015	0.1323	0.0083	0.1868	0.0387	0.0149	0.0000	0.0000	0.0000	0.0011	0.0267	0.1677																															

4

3.2.4 Біграми без пробілів без та без перетинів

Bigrams without spaces non-intersected ---																																																							
	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	ъ	ы	ь	э	ю	я	б																						
а	0.0580	0.1557	0.5825	0.1173	0.3526	0.2346	0.1539	0.5324	0.1628	0.0597	0.6299	1.1968	0.3520	0.7724	0.2011	0.2486	0.3220	0.6340	0.6020	0.0605	0.0360	0.1158	0.0162	0.1613	0.0794	0.0259	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000																							
б	0.1026	0.0001	0.0106	0.0007	0.0041	0.2316	0.0006	0.0004	0.0999	0.0000	0.0282	0.0958	0.0067	0.0368	0.2503	0.0007	0.1244	0.0390	0.0133	0.1246	0.0001	0.0062	0.0011	0.0009	0.0009	0.0255	0.0140	0.4862	0.0009	0.0094	0.0007	0.0584																							
в	0.6400	0.0177	0.0327	0.0326	0.0738	0.5187	0.0071	0.0737	0.5693	0.0000	0.0721	0.1023	0.0410	0.1968	0.8737	0.0873	0.1474	0.0410	0.0717	0.0894	0.0020	0.0098	0.0050	0.0252	0.1319	0.0009	0.0003	0.2866	0.0248	0.0293	0.0007	0.0377																							
г	0.0965	0.0037	0.0118	0.0016	0.1436	0.0459	0.0011	0.0265	0.0964	0.0000	0.0186	0.1777	0.0024	0.0308	0.1013	0.0118	0.0866	0.0128	0.0031	0.0734	0.0044	0.0003	0.0000	0.0047	0.0009	0.0000	0.0000	0.0000	0.0000	0.0003	0.0003	0.0000																							
д	0.4957	0.0062	0.0978	0.0031	0.0078	0.5439	0.0036	0.0033	0.2487	0.0000	0.0316	0.0944	0.0094	0.1988	0.4602	0.0147	0.1837	0.0526	0.0269	0.2129	0.0007	0.0008	0.0204	0.0647	0.0122	0.0110	0.0688	0.0662	0.0937	0.0028	0.0011	0.0472																							
е	0.0316	0.2650	0.4996	0.4740	0.4043	0.3003	0.0308	0.1186	0.2195	0.1665	0.3023	0.2977	0.7347	0.1042	0.1838	0.3710	0.6827	0.6791	0.6289	0.0700	0.0072	0.1268	0.0208	0.1733	0.1156	0.0730	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000																							
ж	0.1669	0.0005	0.0020	0.0009	0.0907	0.4797	0.0017	0.0007	0.1756	0.0000	0.0122	0.0007	0.0105	0.1239	0.0077	0.0030	0.0001	0.0108	0.0028	0.0283	0.0000	0.0000	0.0000	0.0035	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000																							
з	0.6438	0.0234	0.1035	0.0673	0.1025	0.0291	0.0122	0.0079	0.0412	0.0000	0.0201	0.0271	0.0404	0.2395	0.0441	0.0136	0.0259	0.0194	0.0079	0.0336	0.0003	0.0006	0.0011	0.0084	0.0168	0.0000	0.0011	0.0492	0.0190	0.0043	0.0011	0.0578																							
и	0.0356	0.1338	0.5442	0.0950	0.2715	0.3506	0.0565	0.2737	0.1949	0.1590	0.2929	0.5656	0.3833	0.6984	0.1932	0.2296	0.1139	0.0463	0.5569	0.0619	0.0094	0.1767	0.1022	0.2877	0.0577	0.0190	0.0000	0.0000	0.0000	0.0000	0.0274	0.0288																							
й	0.0412	0.0093	0.0719	0.0210	0.0713	0.0126	0.0177	0.0166	0.0792	0.0000	0.0567	0.0276	0.0349	0.0979	0.0564	0.0702	0.0316	0.1467	0.0581	0.0173	0.0044	0.0041	0.0018	0.0065	0.0339	0.0095	0.0011	0.0000	0.0000	0.0000	0.0000	0.0000																							
к	0.9328	0.0395	0.0567	0.0095	0.0234	0.0733	0.0254	0.0099	0.3764	0.0000	0.0306	0.0516	0.0208	0.1448	0.9784	0.0366	0.1740	0.1620	0.0730	0.1814	0.0009	0.001																																	

4 Ентропія та надлишковість

4.1 Біграми з пробілами та з перетинами

— Entropy: 3.945625089661013, Redundancy: 0.8804356033436057 —

4.2 Біграми без пробілів та з перетинами

— Entropy: 4.123476319796595, Redundancy: 0.8750461721273759 —

4.3 Біграми з пробілами та без перетинів

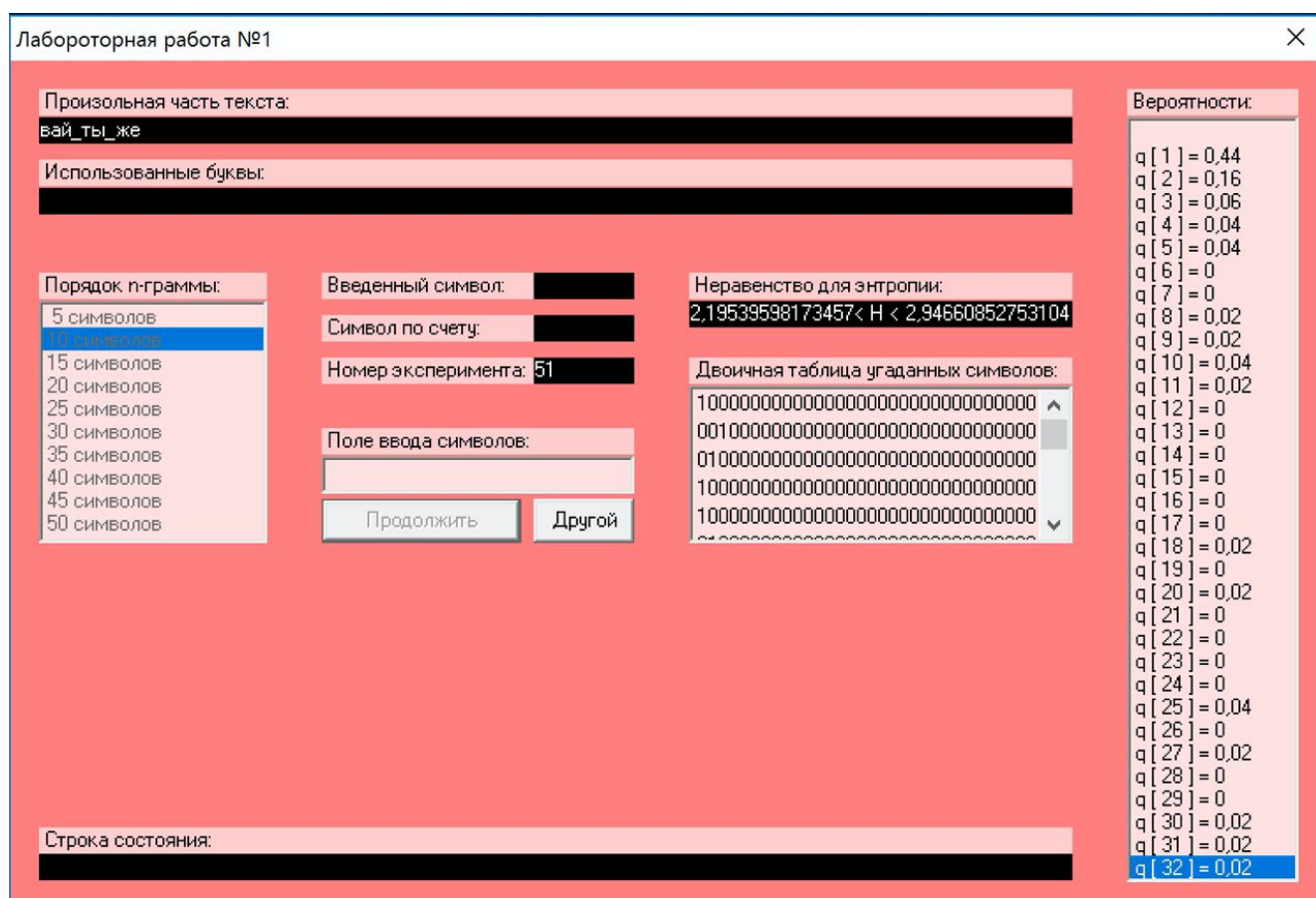
— Entropy: 3.9449776487612955, Redundancy: 0.8804552227648093 —

4.4 Біграми без пробілів та без перетинів

— Entropy: 4.122368938570627, Redundancy: 0.8750797291342234 —

5 CoolPinkProgram.exe

5.1 H10



$$0.9079 < R < 0.9313$$

5.2 H20

Лабораторная работа №1

$$0.9254 < R < 0.9490$$

5.3 H30

$$0.9416 < R < 0.9651$$

6 Висновки

Впродовж цієї лабораторної роботи ми засвоїли що таке ентропія символів джерела, надлишковість символів. Також ми порівняли різні моделі джерела відкритого тексту для наближеного визначення ентропії, здобули практичні вміння та навички необхідних для досить точної оцінки ентропії.