

КРИПТОГРАФІЯ

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого
тексту

Роботу виконали:

студенти групи ФБ-14

Антонова Олександра і Веденкін Артем

Київ-2023

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

Обрали текст: всі томи “Війна і мир” 🤔, автор Лев Толстой, текст написаний російською мовою.

Так як спочатку обраний текст містить багато символів, окрім власне літер, то ми першочергово проводимо фільтрацію тексту та створюємо два файли: з пробілами і без пробілів.

Створюємо функцію для підрахунку частоти букв у тексті, а також обчислення ентропії та надлишку. Результати частоти букв в тексті з пробілами і без пробілів експортуємо у файли

«frequency_of_letters_with_spaces.xlsx» і

«frequency_of_letters_without_spaces.xlsx» відповідно.

Ентропія:

Надлишковість:

H2: 3.976447690967201
R: 0.21838295137504837

Частота біграм з перетином:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	ac	ad	ae	af	ag	ah
a	0.013333	0.001271	0.000988	3.86e-05	1.2e-05	0	0.000116	0.000919	0	6.45e-06	0	0.000257	0.000004	0.000768	1.9e-05	4.6e-06	0.005844	0.000314	4.3e-06	0.006412	6.35e-05	0	0.000339	0.003289	0.000845	0	0.211e-07	0	0.002947	0	0	0	0.948e-05
b	0.002334	0.01671	0.000294	6.88e-06	0.000149	0	3.07e-07	0.000061	0	6.14e-07	0	0	5.52e-06	0.000983	0.000078	0.000416	0.000004	0.003252	0.000281	0.000470	0.000084	0.000227	0	0.001302	2.09e-05	0	0	0.000211	0	0	0	0	0.212e-05
c	0.008726	1.13e-05	0.000179	0.000098	2.48e-06	9.21e-07	0.000162	0.000057	3.07e-07	0	0	3.07e-07	3.98e-05	3.99e-06	0.000216	0.000019	6.14e-07	0.000209	0.000191	0.000161	0	0.577e-05	0.004766	0	0	0.000095	3.14e-07	6.14e-07	3.07e-07	3.07e-07	0.001195	0	3.41e-05
d	0.016561	0.000581	0.000923	0.00709	0.000195	2.78e-06	0.000361	0	0	3.07e-07	3.44e-05	0.001676	9.21e-07	3.07e-06	0.000449	0.0001185	4.57e-05	1.26e-05	0.000204	0.000291	0.000264	0.000391	6.35e-05	1.01e-05	0.000146	0	0	1.23e-06	6.14e-07	0.000169	4.67e-05	0	0.000198
e	0.005048	0.000374	9.32e-06	0.000536	0.000189	0	3.99e-05	0	0	0	1.15e-05	0.001106	0.000989	1.84e-06	0.001378	4.21e-05	0.000064	6.26e-05	0.00122	0.002844	0.000601	0.000898	0.00023	0.000748	0	0	0.000231	0	0	3.07e-07	6.14e-07	0.001208	0
f	0.000975	0.000791	7.95e-05	1.47e-05	0.000075	0.001905	0.000308	0.000154	0	0	0	9.32e-06	0.001109	4.14e-06	2.46e-06	9.21e-07	0.001108	5.11e-05	0.000264	0.000953	4.87e-05	0.001381	0.000845	2.46e-05	0.001132	4.3e-06	0	0.000118	0.001341	0	0	6.14e-07	
g	0.000363	0.001732	9.32e-06	0.84e-05	0	0.000403	0.015588	0.000364	0	0	0	0	0.000211	0.000162	2.71e-05	9.21e-07	7.06e-06	0.002185	0.000429	0.000681	0.000892	6.42e-05	0.002746	0.000905	0.000892	6.42e-05	5.22e-06	0	0	0.001742	0	0.133e-06	
h	6.14e-07	0.000994	2.06e-05	3.78e-05	3.9e-05	3.9e-05	3.1e-05	0.000556	0	0	3.9e-05	1.33e-05	0.000519	0.000452	5.84e-05	1.14e-05	5.22e-06	3.71e-06	0.002068	0.00042	7.37e-06	2e-05	0.001595	0.014108	0.000272	2.46e-06	8.35e-05	0.000118	0	3.07e-07	0.000692	0	9.21e-07
i	0.00234	0	0.001989	0.000274	4.02e-05	0.000881	0.00201	0	1.26e-05	0	3.94e-05	0.000288	0	2.78e-06	9.98e-05	1.78e-05	3.07e-07	1.13e-06	8.9e-06	6.08e-05	3.96e-05	0.000349	0.000297	6.34e-05	1.73e-05	1.23e-06	4.51e-05	3.78e-06	0.000217	0	3.07e-07	0.000126	0
j	0.001842	3.68e-06	3.07e-07	0.004133	0.000396	7.87e-06	0.0001462	0	0	0.000185	0	6.32e-05	3.07e-07	0	3.07e-07	8.8e-06	0.00011	2.34e-05	6.75e-06	0.000316	0.000619	0.000517	0.000867	0.000483	0.000633	7.21e-05	0.000104	1.84e-06	3.6e-05	0.000367	0	0	0.000198
k	0.000558	0.000114	3.07e-07	1.23e-06	0.000557	2.7e-05	0.000177	9.21e-07	0	0.001265	0.000189	0.000486	0.000346	0	0	6.45e-06	2.96e-05	0.000686	0.00025	1.81e-05	0.002674	0.01372	0.000904	0.000599	6.14e-07	0.002009	5.53e-06	4.3e-06	0.000392	0	0	0	0
l	0.011706	0.000371	6.14e-06	4.14e-07	0	0.000418	0.000472	1.23e-06	0	0.568e-05	0.000789	0.017787	0.000311	0	0	2.15e-06	4.45e-06	0.000203	6.82e-05	0	1.66e-05	0.00074	0.000111	0.001321	0.000546	7.86e-06	0.000344	0.000491	1.41e-06	0	0	0	3.07e-07
m	0	0.000126	6.75e-06	9.21e-07	3.07e-07	1.23e-06	0.00151	9.21e-07	4.91e-06	0	0.00016	8.84e-05	0.000384	1.84e-06	0	9.21e-07	0	4.91e-06	2.15e-05	0.001289	1.26e-05	9.21e-05	1.29e-05	0.001281	1.26e-05	9.76e-05	9.21e-07	6.83e-05	0.000346	0	0	0.000109	
n	1.29e-05	0	0.000886	0.000462	1.23e-06	1.33e-06	2.01e-06	0	0.568e-06	1.11e-05	0.000701	0.000421	5.33e-06	0.000862	3.07e-07	5.53e-06	0.000346	0	0.000106	0.00101	7.95e-05	7.77e-05	0.000108	7.98e-06	6.14e-06	1.23e-05	8.35e-05	0	0.000204	2.78e-05	0	0	9.21e-07
o	0.000958	0.000611	0.000383	0.000384	1.23e-06	1.33e-05	1.13e-06	0	0	0.000986	0.000387	0.000336	0.000445	2.46e-06	0.000429	0	0.000999	3.97e-07	6.14e-06	0.00011	4.48e-06	4.66e-05	2.49e-05	1.2e-05	1.14e-05	7.46e-05	1.33e-05	1.23e-06	2.06e-05	3.16e-05	3.07e-07	0	0.000208
p	0.000457	0.000381	0	0.00239	0.000688	3.07e-06	0.000962	0	0	0.000463	0.001556	0.000487	0.001022	7.98e-06	0.000769	0.00086	0.001702	2.92e-05	0.000993	0	1.13e-06	0.000266	0.000311	0.000102	0.000305	2.76e-06	0.000108	1.53e-06	2.46e-06	2.15e-06	0	0	0.000101
q	0.000536	0.000491	0.00021	3.07e-07	6.14e-07	0.000231	0.001172	3.07e-07	3.07e-07	3.07e-07	0.000226	0.001498	8.5e-05	1.14e-05	1.21e-06	0.000615	0.000941	0.000412	0.000311	0	4.57e-05	7.98e-05	0.00014	7.37e-06	0	0.214e-07	0.000545	0.000344	1.23e-06	2.46e-06	0	0	6.15e-05
r	0.014572	0.000565	0.000731	0.000141	0	0.000567	0.000178	6.14e-07	0	3.07e-07	0	0.002685	9.92e-05	8.9e-06	0.000341	2.24e-05	0.000262	0.000315	0.000433	0.000776	0.00084	0	3.07e-07	9.21e-07	0.00073	0	3.15e-06	0.000564	0.00056	0	0.133e-05	0	0.133e-05
s	0.011162	0.00473	3.78e-05	0.001007	7.34e-05	0	0	3.07e-06	1.78e-05	0	0	0.001029	0.000148	7.97e-06	0.000117	1.57e-05	4.41e-05	8.9e-06	9.21e-05	0.000217	0.000127	0.000254	4.51e-05	0.00025	0	1.53e-06	0	0	0.000281	0	3.07e-07	1.84e-06	0
t	0.014847	2.46e-05	0.000239	0.001002	0.001599	0.000377	0.000239	0	0	0	0.000144	2.78e-06	0	0.491e-06	7.37e-06	0.000117	1.57e-05	4.41e-05	8.9e-06	9.21e-05	0.000217	0.000127	0.000254	4.51e-05	0.00025	0	1.53e-06	0	0	0.447e-05	0	0.447e-05	0
u	0.000466	0.000892	0.000296	0.001087	3.98e-06	0.000742	0.001036	0	0	0.001264	6.85e-05	0.000366	1.23e-06	0	0.000516	0.000235	5.56e-05	7.76e-05	0.000118	0.000274	0.000941	0.000107	0.004982	0	0	7.67e-06	3.07e-07	0	0	0	0	3.07e-07	7.15e-05
v	0.016389	0.0002831	0	0.000713	0.000232	0.000119	0.000568	0	1.07e-05	0	0.000395	0.000743	3.98e-05	1.84e-06	9.21e-07	0.000458	5.83e-06	0.000117	0.000769	6.14e-07	0.000864	0.002757	0.00284	4.51e-05	0	0	0	9.21e-07	1.53e-06	0	0	6.14e-07	
w	0.00671	0.000463	0.001067	0.000226	0.000364	0.001547	0.00371	1.53e-06	4.91e-05	0	0.00025	0.000876	1.23e-05	3.07e-07	0	4.3e-06	0.003328	0.000883	0.000469	3.07e-07	0.000123	0.00138	0.000466	0.000488	0.000912	3.07e-06	1.46e-05	0	8.35e-05	0	0	4.8e-06	
x	0.000347	0.000453	7.18e-05	0.000186	0.000149	0.000781	6.14e-07	8.81e-05	7.67e-05	0	0	0.00019	3.07e-07	3.07e-07	0.000855	0.000416	3.68e-06	0.000286	3.03e-05	0.000517	0.000542	6.51e-05	2.15e-06	0	0	1.84e-06	0	0	0	0	0	0	
y	0.000827	0.000433	6.45e-05	0.000208	0.00112	9.48e-05	0.000156	1.84e-06	4.5e-06	3.99e-06	0.000116	0.001299	0	4.91e-05	9.52e-06	0.000398	2.15e-06	3.07e-07	0.000243	0	0.000395	0.001705	2.76e-05	2.21e-05	0.000171	0	0	0	0	0	0	0	
z	0.001032	0.000403	0.000087	0.000087	0.000107	0.001795	0.000837	0	2.7e-05	6.45e-06	0.000106	0.001448	0.001133	0	0.002459	4.51e-06	3.07e-07	1.15e-06	0	0.000286	0.000134	1.93e-05	0.000345	0.000544	0.000107	0.000361	0	0.000165	0	0	0	0	0
aa	0.000323	0.000908	0	0.000768	0.000196	0.000375	0.0006302	0	0	0.000192	0.000247	0.000497	0.000173	3.07e-07	0	0.000584	0.000218	1.36e-05	0.000792	3.07e-07	0.000774	0.000225	3.07e-06	0	0.000024	0	0	0.000044	0	0	0.000431	0	
ab	0.00012	7.15e-05	8.9e-05	0	0.000602	0.000199	0.000431	5.53e-06	2.75e-05	4.85e-05	0.001524	0.00277	0.0002	6.32e-05	0	0.000049	0.000394	1.55e-05	0.001609	6.14e-07	3.07e-07	0.002947	0.00035	0.000113	0	9.21e-07	0.000032	0.000318	6.14e-07	1.41e-05	0	0.798e-06	
ac	0.000727	0.00077	2.78e-06	1.71e-05	3.07e-07	0.00013	0.000353	9.21e-07	1.69e-05	0	0.000431	0.000962	0.001195	7.55e-05	0.00015	0.000735	0.000988	0	9.21e-07	0.000288	0.000413	0.001589	3.07e-07	0.000138	0	0	0.000221	8.2e-05	0	0	0	0	
ad	5.62e-05	0.000197	8.81e-05	0.000185	0.000089	2.82e-05	6.16e-06	0.000893	0.001959	0.000279	0.000624	0.000628	0.000399	0.001199	0.0006424	2.23e-06	0.000442	0.000551	0.000458	0.000248	3.71e-06	0.000208	9.64e-06	5.19e-06	7.42e-06	0	0	0.000436	0.000662	0.01202	0	0	
ae	0.001698	0.000398	0.000188	0.000088	0.000022	0.000208	2.77e-05	7.42e-07	0.000417	0.000288	0.0003771	0.001339	0.000123	0.0007	0.0005462	0.000445	0.010187	0.003141	0.002946	0.00899	0.000127	0.000332	0.000347	0.0001									

Частота біграм з перетином:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	
a	0.00077	0.00119	0.00592	0.00182	0.00491	0.00021	2.8E-05	3.71E-07	0.00155	0.00080	0.00339	5.3E-05	0.00134	0.00045	0.00077	0.00027	0.00204	0.00026	0.00136	0.00079	0.00178	0.00013	0.00046	0.00071	0.00097	0.00065	0.00046	0.00087	0.00049	0	0.00045	0.00119	0.00029	
b	0.00146	1.6E-05	0.00027	7.98E-05	1.98E-05	0.00018	5.9E-05	3.71E-07	7.79E-05	0.00023	0.00142	0.00069	0.00033	0.00043	0.00002	0.00042	0.00010	0.00019	1.48E-06	0.00046	0.00019	0.00014	0.00087	9.27E-06	0.00015	1E-05	1.15E-05	1.04E-05	3.71E-07	0	0.00047	0.00063	0.00051	
v	0.00097	0.00010	0.00083	0.00028	0.00119	0.00068	5.73E-05	3.71E-06	2.21E-05	0.00101	0.00142	0.00167	0.00074	0.00078	0.00081	0.00081	0.00055	0.01209	3.71E-06	0.00099	0.00028	0.00365	0.00154	2.15E-05	0.00019	5.27E-05	4.08E-05	2.78E-05	1.81E-06	3.71E-07	0.00173	0.00024	0.00167	
t	0.00141	1.34E-05	0.00042	4.49E-05	0.00106	0.00182	1.11E-05	3.93E-06	1.89E-05	0.00012	0.00127	0.00028	0.00012	0.00039	0.00079	0.00077	0.00088	5.97E-06	0.0004	0.00016	0.00015	0.00015	4.45E-06	0.00019	1.89E-05	0.00019	1.89E-05	0.00019	2.97E-06	0	0.00027	0.00117	0.00059	
d	0.00128	2.78E-05	0.00041	0.00126	1.53E-05	0.00082	2.89E-05	7.05E-06	0.00063	0.00197	0.00168	0.00034	0.00068	0.00043	0.00026	0.00039	0.00016	0.00048	0.00016	0.00048	0.00016	0.00048	0.00016	0.00027	9.44E-06	8.33E-06	6.31E-06	7.42E-07	0	0.00048	0.00062	0.00127		
e	0.00166	0.00142	0.00172	0.0007	0.00076	0.00127	1.97E-05	3.71E-07	0.00193	0.00036	0.00072	0.00137	0.00017	0.00071	0.00035	0.00049	0.01017	0.00033	0.00039	0.00085	0.00016	0.00047	0.00021	0.00044	0.00016	0.00012	0.00063	0.00144	0.00025	0.00196	8.36E-05	0.00047		
k	0	3.71E-07	1.48E-06	0	1.48E-06	2.4E-06	0	0	3.71E-07	0	0	0	0	1.11E-06	5.39E-06	7.42E-07	3.71E-07	3.71E-07	0	3.71E-07	0.00012	4.45E-06	0	0	0	0	3.71E-07	7.42E-07	7.42E-07	7.42E-07	0	0	0	0
n	0.00038	4.08E-05	0.00028	3.12E-05	2.6E-05	0.00046	6.43E-05	7.42E-07	3.8E-05	4.15E-05	0.00025	5.49E-05	6.75E-05	0.00010	0.00016	7.72E-05	4.75E-05	0.00078	2.23E-06	2.19E-05	5.82E-05	0.00019	0.00010	3.71E-07	3.89E-05	4.82E-06	2.6E-06	2.97E-06	3.71E-07	0	0.00017	4.3E-05	0.00161	
m	0.00149	8.33E-06	5.9E-05	1.74E-05	2.41E-05	0.00137	4.82E-06	0	2.26E-05	0.00018	0.00029	6.27E-05	0.00033	0.00013	0.00043	0.00017	5.3E-05	0.00148	3.71E-07	0.00079	0.00013	0.00011	0.00011	1.85E-06	4.71E-05	1.48E-06	8.89E-06	1.48E-06	0	0	7.01E-05	8.9E-05	0.00084	
y	0.00042	1.15E-05	0.0007	6.51E-05	4.97E-05	0.00188	2.41E-05	3.71E-06	1.6E-06	0.00015	0.00061	0.00025	0.00156	3.72E-05	0.00015	0.00010	0.00019	0.00015	0.00121	7.42E-07	7.84E-05	0.00018	0.00013	0.00118	5.19E-06	8.2E-05	5.66E-06	3.71E-06	2.23E-06	7.42E-07	0	0.00049	9.2E-05	0.00142
z	0.00125	0.00087	0.00494	0.00098	0.00291	0.00015	1.52E-05	3.71E-07	0.00158	0.00045	0.00174	0.00047	0.00016	0.00024	0.00079	0.00047	0.00021	0.00087	0.00097	0.00176	0.00042	0.00065	0.00074	0.00054	0.00033	0.00163	0.00111	0.00087	3.71E-07	0.00087	0.00033	0.00061		
h	0	0.00439	0.00388	7.42E-07	0.00085	0	0	0	0	0	0.00041	0	0	0	0.00118	0.00012	0.00088	0	0.00027	0.00173	0.00041	0.00157	0	1.22E-05	0	0.00016	0	3.71E-07	0	0	0	0		
g	0.00078	0	3.71E-07	3.71E-07	3.71E-07	0.00070	0	1.1E-05	0	0	0.00137	0.00158	2.4E-06	3.71E-07	1.48E-06	7.42E-07	0	0.00082	0	3.71E-07	0	0	0	7.42E-07	7.75E-05	0	3.71E-07	0	0	0	0	1.11E-06	0	
l	0.00086	0.00024	0.00083	0.00017	0.00012	0.00219	2.3E-05	5.97E-05	0.00010	0.00027	0.00034	0.00044	0.00066	0.00072	0.00077	0.00068	0.00043	0.00479	8.09E-05	0.00064	0.00091	0.00073	0.00195	1.28E-05	0	0.00021	0.00016	0.00073	0.00045	1.48E-06	3.71E-07	0.00191	0.00021	0.00099
x	0.00124	0.00089	0.00191	0.00198	0.00011	0.00152	1.11E-05	0.00017	1.61E-06	0.00026	0.00043	0.00072	0.00026	0.00062	0.00018	0.00024	4.56E-05	0.00025	0.00024	0.00116	0.00028	0.00052	0.00146	8.01E-05	0.00027	6.51E-06	4.53E-05	0.00047	0	0	0.00124	0.00033	0.00122	
o	0.00047	4.8E-05	0.00123	4.47E-05	0.00018	0.00449	9.44E-06	7.05E-06	1.19E-05	0.00038	0.00079	0.00018	0.00039	0.00013	0.00017	0.00021	0.00015	0.00045	1.89E-06	0.00089	0.00117	0.00020	0.00134	7.42E-06	0.00017	1.11E-05	1.04E-05	2.38E-05	3.71E-07	0	0.00018	0.00162	0.00078	
u	0.00073	0.00032	0.00188	0.00047	0.00102	0.00181	2.89E-05	3.8E-05	0.00137	0.00194	0.00154	0.00041	0.00083	0.00105	0.00105	0.00105	0.00105	0.00083	0.00083	8.23E-05	0.00081	0.00142	0.00188	0.00106	1.48E-05	0.00042	2.82E-05	0.00008	0.00045	3.45E-05	0	0.00029	0.00033	0.00157
c	0.00143	0.00284	0.00011	0.00284	0.00048	0.00194	3.45E-05	3.71E-07	7.01E-05	0.00082	0.00188	0.00021	0.00027	0.00087	0.00072	0.00047	0.01032	0.00033	0.00094	0.00026	0.00033	0.01142	0.00044	0.00014	0.00028	0.00049	8.89E-05	0.00013	3.71E-07	3.71E-07	0.00079	0.00042	0.00086	
p	0.00069	7.42E-06	0.00102	0.00026	0.00033	0.00027	4.82E-05	3.49E-05	1.3E-05	0.00026	0.00029	0.00062	0.00074	0.00089	0.00062	0.01043	0.00073	0.00079	5.8E-05	0.00074	0.00021	0.00029	0.00136	3E-05	0.00041	3.49E-05	3.8E-05	4.45E-05	7.42E-07	0	0.00095	0.00116	0.00112	
r	0.00098	0.00129	0.00148	0.00197	0.00029	0.00087	4.12E-05	2.04E-05	3.19E-06	0.00024	0.00143	0.00045	0.00052	0.00182	0.00024	0.00024	0.00012	0.00072	0.00122	5.97E-05	0.00047	0.00085	0.00085	0.00028	7.42E-06	4.12E-05	7.05E-06	4.08E-06	0	0.00023	0.00029	0.00032		
s	0.00055	9.53E-05	0.00048	0.00024	0.00059	0.00104	5.89E-05	4.71E-05	4.15E-05	0.00041	0.00054	0.00146	0.00115	0.00089	0.00024	0.00107	0.00183	0.01038	1.37E-05	0.00178	0.00199	0.00194	3.66E-05	0.00033	3.66E-05	2.82E-05	1.84E-06	7.42E-07	3.71E-07	0.00051	0.00011	0.00116		
i	0.00037	8.44E-06	0.00088	4.23E-05	0.00074	0.00045	0.00104	0.00284	1.89E-05	8.75E-05	0.00061	0.0006	0.00027	0.00072	0.00047	0.00014	0.00068	0.00078	4.89E-05	0.00072	0.00082	0.00045	0.00187	2.4E-05	0.00019	1.48E-05	0.00047	0.00011	3.71E-07	0	0.00041	0.00048	0.00113	
q	0.00055	0.00129	0.00188	0.00072	0.00182	0.00081	1.54E-05	0	0.00017	0.00039	0.00052	9.44E-05	0.00014	0.00172	0.00148	0.00084	0.00028	0.00018	0.00078	0.00078	0.00025	0.00174	0.00021	9.44E-05	0.00021	0.00088	0.00088	0.00034	0.00012	3.71E-07	0.00023	0.00017	0.00024	
f	0.00099	0	3.45E-05	1.11E-05	2.23E-05	9.44E-05	7.42E-07	1.85E-06	3.71E-07	7.79E-06	0.00017	2.83E-05	7.68E-05	2.49E-05	4.75E-05	6.05E-05	3.12E-05	0.00003	2.26E-05	3.82E-05	4.71E-05	3.49E-05	4.44E-05	8.82E-06	2.71E-05	1.48E-06	7.42E-07	3.71E-07	0	0	0.00011	4.49E-05	7.01E-05	
w	0.00125	0.00108	8.44E-05	9.44E-06	5.08E-05	0.00129	7.79E-06	7.79E-06	1.11E-06	6.31E-06	0.00017	0.00123	8.88E-05	4.08E-05	3.89E-05	7.05E-05	5.44E-05	0.00072	3.71E-07	0.00013	0.00023	0.00023	0.00043	1.48E-06	1.85E-05	1.48E-06	2.23E-06	7.42E-07	0	0	0.00011	4.49E-05	7.01E-05	
h	0.00136	3.34E-06	4.19E-05	1.59E-06	0.00018	0.00026	1.11E-06	1.11E-06	0	8.59E-06	0.00138	1.52E-05	4.38E-05	2.52E-05	1.08E-05	4.04E-05	0.00081	0.00013	1.52E-05	9.81E-05	0.00014	1.97E-05	3.71E-07	1.19E-05	0	0	0	0	0	0	0.00011	4.49E-05	7.01E-05	
o	0.00132	2.41E-05	0.00016	0.00016	6.9E-05	0.00157	5.58E-05	1.11E-06	4.4E-05	4.01E-05	0.00024	0.00017	0.00015	0.00093	0.00049	0.00034	0.00081	9.74E-05	0.00019	0.00013	0.00013	0.00028	0.00019	2.23E-06	7.57E-05	3.68E-06	7.79E-06	1.85E-06	0	0	0.00048	0.00029	0.00048	
n	0.00158	4.82E-06	0.00179	1.26E-05	0.00017	0.00088	2.23E-06	3.71E-07	3.71E-07	2.34E-05	0.00052	0.00054	0.00019	3.08E-05	5.19E-05	6.44E-05	4.19E-05	0.00116	6.31E-06	0.00033	0.00016	1.6E-05	0.00024	2.23E-06	2.78E-05	1.48E-06	0.00015	2.6E-06	0	0	0.00056	0.00019	6.38E-05	
u	0.00028	0.00021	8.8E-06	3.71E-07	7.42E-07	0.00076	0	0	0	1.11E-06	0.00016	5.93E-06	2.97E-06	2.23E-06	6.31E-06	8.53E-06	0.00013	0.00023	2.23E-06	5.53E-05	1.48E-06	1.78E-05	0.00006	0	1.48E-06	0	0	0	0	0	0	0.11E-05	0.00048	0.00049
z	0	0.00012	1.45E-05	0	0.00019	0	0	0	0	4.71E-05	0.00071	7.42E-07	0	0	0	1.11E-06	0.00012	3.71E-05	0.00177	3.71E-07	0.00084	0.00058	0.00027	0.00019	3.71E-07	3.71E-06	0	0	0.0002	0.00029	7.71E-05	0	0	0
o	0.00048	4.45E-06	4.82E-06	3.34E-06	4.93E-05	0.00043	0	0	4.3E-05	7.64E-05	0.00042	3.71E-07	4.82E-06	6.31E-06	0.00128	3.64E-05	0.00138	0.00041	0	0.00014	0.00011	5.53E-05	0.00103	2.6E-06	7.42E-07	0	0	0						

$$1.50557613404877 < H^{(20)} < 2.13366068968819$$

Лабораторная работа №1

Произвольная часть текста:
нии_чисел_но_понятие_о_дobre_и_эле_не_в_большей_мере_зависит_от_чьего_то_вк

Использованные буквы:
я, ю, к,

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: e

Символ по счету: 4

Номер эксперимента: 50

Неравенство для энтропии:
 $1.50557613404877 < H < 2.13366068968819$

Двоичная таблица угаданных символов:

00000001000000000000000000000000
10000000000000000000000000000000
00000000000000000000000000000000
00000000000000000000000000000000
10000000000000000000000000000000

Поле ввода символов:
e

Продолжить Другой

Вероятности:

q[1] = 0.64
q[2] = 0.06
q[3] = 0.02
q[4] = 0.06
q[5] = 0.04
q[6] = 0
q[7] = 0
q[8] = 0.04
q[9] = 0
q[10] = 0
q[11] = 0.02
q[12] = 0
q[13] = 0
q[14] = 0
q[15] = 0
q[16] = 0
q[17] = 0
q[18] = 0
q[19] = 0.02
q[20] = 0.02
q[21] = 0
q[22] = 0
q[23] = 0.02
q[24] = 0
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0.04
q[29] = 0.02
q[30] = 0
q[31] = 0
q[32] = 0

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

$$1.69546356271206 < H^{(30)} < 2.35603171251419$$

Лабораторная работа №1

Произвольная часть текста:
ьшей_мере_зависит_от_чьего_то_вкуса_и_мнения_чем_таблица_умножения_а_теперь

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: _ (пробел)

Символ по счету: 1

Номер эксперимента: 50

Неравенство для энтропии:
 $1.69546356271206 < H < 2.35603171251419$

Двоичная таблица угаданных символов:

10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
00010000000000000000000000000000
10000000000000000000000000000000

Поле ввода символов:

Продолжить Другой

Вероятности:

q[1] = 0.58
q[2] = 0.12
q[3] = 0.02
q[4] = 0.04
q[5] = 0.04
q[6] = 0
q[7] = 0.02
q[8] = 0
q[9] = 0
q[10] = 0
q[11] = 0.02
q[12] = 0
q[13] = 0
q[14] = 0
q[15] = 0
q[16] = 0
q[17] = 0
q[18] = 0.04
q[19] = 0
q[20] = 0.02
q[21] = 0
q[22] = 0
q[23] = 0.04
q[24] = 0
q[25] = 0.02
q[26] = 0.02
q[27] = 0
q[28] = 0
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0.02

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

Обчислимо надлишковість для $H^{(10)}$, $H^{(20)}$, $H^{(30)}$:

- для $H^{(10)}$:
 1. $R = 0.6018914791415$
 2. $R = 0.455568211464486$
- для $H^{(20)}$:
 1. $R = 0.698884773190246$
 2. $R = 0.573267862062362$
- для $H^{(30)}$:
 1. $R = 0.660907287457588$
 2. $R = 0.528793657497162$

Надлишковість обчислюємо за формулою $R = 1 - \frac{H_{\infty}}{H_0}$, $H_0 = \log_2 32 = 5$.

Висновки:

Під час виконання комп'ютерного практикуму ми ознайомились з поняттям ентропії на символ джерела та його надлишковості, написали програми для підрахунку частот букв і частот біграм в тексті, попрацювали з CoolPinkProgram, де оцінювали та порівнювали $H^{(10)}$, $H^{(20)}$, $H^{(30)}$. В результаті експериментів ми отримали значення ентропії та надлишковості власного джерела, які наближаються до значень, отриманих з прикладеної програми.