

## 1 Определение количества информации

Использованный текст: Толкин Джон Рональд Руэл «Властелин Колец. Две Крупости» [1].

Создана программа на языке программирования Kotlin для расчёта значений необходимых для данной лабораторной работы. Исходный код программы представлен в приложении А.

Таблица расчёта энтропии источника представлена в приложении А.

Таблица 1 — Расчёт энтропии

№	Символ	Код символа	Число вхожде- ния в текст	Вероятность	I
1	0	48	7	1.2207265E-5	16.321901
2	1	49	16	2.790232E-5	15.129255
3	2	50	7	1.2207265E-5	16.321901
4	3	51	8	1.395116E-5	16.129255
5	4	52	4	6.97558E-6	17.129255
6	5	53	4	6.97558E-6	17.129255
7	6	54	4	6.97558E-6	17.129255
8	7	55	5	8.719475E-6	16.807327
9	8	56	4	6.97558E-6	17.129255
10	9	57	4	6.97558E-6	17.129255
11	а	1072	35026	0.061081666	4.033117
12	б	1073	7616	0.013281505	6.2344375
13	в	1074	18846	0.032865446	4.9272847
14	г	1075	9707	0.01692799	5.8844457
15	д	1076	15221	0.026543826	5.23548
16	е	1077	36376	0.06343593	3.9785562
17	ё	1105	314	5.4758304E-4	10.834635
18	ж	1078	4177	0.0072842496	7.101004
19	з	1079	7888	0.013755844	6.1838117
20	и	1080	30442	0.05308765	4.23548
21	й	1081	4454	0.0077673085	7.0083694
22	к	1082	13740	0.023961117	5.383161
23	л	1083	24790	0.04323116	4.5317845

№	Символ	Код символа	Число вхожде- ния в текст	Вероятность	I
24	м	1084	15943	0.027802918	5.16862
25	н	1085	30425	0.053058006	4.2362857
26	о	1086	51905	0.09051687	3.4656694
27	п	1087	12762	0.022255588	5.4896884
28	р	1088	21819	0.038050048	4.715958
29	с	1089	24347	0.042458612	4.557799
30	т	1090	25912	0.04518781	4.4679227
31	у	1091	13279	0.023157183	5.4323964
32	ф	1092	1339	0.0023350755	8.742315
33	х	1093	4985	0.008693317	6.8458776
34	ц	1094	961	0.0016758832	9.220862
35	ч	1095	5620	0.00980069	6.672901
36	ш	1096	4044	0.0070523117	7.147688
37	щ	1097	1511	0.0026350254	8.567967
38	ъ	1098	57	9.9402016E-5	13.296366
39	ы	1099	9145	0.015947921	5.9704876
40	ь	1100	9525	0.0166106	5.911752
41	э	1101	2219	0.003869703	8.013561
42	ю	1102	1986	0.0034633756	8.173605
43	я	1103	9238	0.016110102	5.9558907
44		32	93386	0.16285539	2.6183367
45	.	46	8230	0.014352256	6.1225786
46	,	44	11362	0.019814136	5.657326
47	:	58	372	6.48729E-4	10.590096
48	;	59	104	1.8136509E-4	12.428816
49	-	45	4286	0.0074743344	7.063839
50	(	40	7	1.2207265E-5	16.321901
		Всего симво- лов в тексте:	50		
			Полная вероят- ность:	1.0	
				Энтропия ис- точника:	4.5080266

Формулы использованные для расчёта:

- вероятность вхождения символа в текст по формуле (1);

$$p_i = \frac{k_i}{K} \quad (1)$$

где:

$k_i$  — количество символов в тексте,

$K$  — количество всех символов;

- Столбец энтропии для каждого символа по формуле (2)

$$I_i = -\log_2 p_i \quad p_i \neq 0 \quad (2)$$

- Средняя энтропия по формуле

$$I_{cp} = \sum p_i \cdot I_i \quad (3)$$

На рисунке представлен рисунок из созданной программы:

№	Символ	Код символа	Число вхождений	Вероятность	I
1	0	48	7	1.2207265E-5	16.321901
2	1	49	16	2.790232E-5	15.129255
3	2	50	7	1.2207265E-5	16.321901
4	3	51	8	1.395116E-5	16.129255
5	4	52	4	6.97558E-6	17.129255
6	5	53	4	6.97558E-6	17.129255
7	6	54	4	6.97558E-6	17.129255
8	7	55	5	8.719475E-6	16.807327
9	8	56	4	6.97558E-6	17.129255
10	9	57	4	6.97558E-6	17.129255
11	a	1072	35026	0.061081666	4.033117
12	б	1073	7616	0.013281505	6.2344375
13	в	1074	18846	0.032865446	4.9272847
14	г	1075	9707	0.01692799	5.8844457
15	д	1076	15221	0.026543826	5.23548
16	е	1077	36376	0.06343593	3.9785562
17	ё	1105	314	5.4758304E-4	10.834635
18	ж	1078	4177	0.0072842496	7.101004
19	з	1079	7888	0.013755844	6.1838117
			573429	1.0	4.5080266

Рисунок 1 — Таблица энтропии из программы

Для таблицы 2

## **СПИСОК ИСТОЧНИКОВ**

1. Джон Рональд Руэл Толкин. Властелин Колец Две Крепости: ISBN:978–5–17–89238–9. АСТ. 384 с.