

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	3
1 Определение количества информации .....	4
1.1 Вывод по сравнению кодировок .....	8
2 Кодирование дискретных источников информации методом Шеннона-Фано	9
2.1 Пример декодирования сообщения .....	11
2.2 Вывод .....	14
3 Кодирование дискретных источников информации методом Д.Хаффмана	14
3.1 Пример декодирования сообщения .....	17
3.2 Вывод .....	19
ЗАКЛЮЧЕНИЕ .....	21

## **ВВЕДЕНИЕ**

В данном отчёте представлено выполнение лабораторной работы №7. Основной целью данной работы являются получение практических навыков численного определения количества информации, содержащегося в сообщении, а также освоение методов построения кодов дискретного источника информации, используя конструктивный метод, предложенный К. Шенноном и Н. Фано, и метод Хаффмана. Также в рамках работы на примере демонстрируется однозначность раскодирования имеющегося сообщения.

Лабораторная работа состоит из 3 частей, выполняемых последовательно и оцениваемых отдельно.

## 1 Определение количества информации

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

Создана программа на языке программирования Kotlin с использованием фреймворка Compose Multiplatform для расчёта значений необходимых для данной лабораторной работы. Исходный код программы прикреплён к отчёту.

Таблица расчёта энтропии источника представлена в таблице 1.

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
24	м	1084	15943	0.027802918	5.16862
25	н	1085	30425	0.053058006	4.2362857

№	Символ	Код символа	Число вхожде- ния в текст	Вероятность	I
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 символов 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;

— количество информации для каждого символа по формуле (2)

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

— Энтропия источника по формуле

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

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

В таблице 2 представлено сравнение кодировки ASCII и использование равномерного кода на основе Хартли.

Таблица 2 — Расчёт энтропии

	Неопределённость	Разрядность кода	Абсолютная избыточность	Относительная избыточность
При кодировании сообщения стандартной кодо-	8 бит	8 бит	3.4919734	1.3559962

	Неопределён- ность	Разрядность кода	Абсолютная избыточность	Относитель- ная избыточ- ность
вой таблицей ASCII				
При исполь- зовании рав- номерного кода, постро- енного на ос- нове меры Хартли	5.643856 бит	6 бит	1.1358294	0.20125061

Для расчёта данной таблицы 2 использовались следующие формулы:

- Неопределённость (или энтропия) равномерного источника:

$$H_{\max} = \log_2 A \quad (4)$$

где:

$A$  – количество символов в алфавите

Для ASCII-кода:  $\log_2 128 = 7$  бит

Для равномерного кода на основе меры Хартли:

$$\log_2 50 \approx 5.64$$

- Разрядность кода по формуле (5)

$$n = \lceil \log_2 A \rceil = \lceil H_{\max} \rceil \quad (5)$$

Для ASCII-кода: 8 бит

Для равномерного кода на основе меры Хартли:  $\lceil 5.64 \rceil = 6$

- Абсолютная избыточность по формуле (6)

$$D_{\text{abc}} = H_{\max} - H \quad (6)$$

Для ASCII-кода:  $7 - 4.5080266 \approx 2.49$

Для равномерного кода Хартли:

$$5.643856 - 4.5080266 \approx 1.14$$

- Относительная избыточность по формуле (7)

$$D = \frac{D_{\text{abc}}}{H_{\max}} = \frac{H_{\max} - H}{H_{\max}} \quad (7)$$

Для ASCII-кода:  $\frac{2.49}{7} \approx 0.36$

Для равномерного кода Хартли:  $\frac{1.14}{5.64} \approx 0.2$

### **1.1 Вывод по сравнению кодировок**

Максимальная энтропия  $H_{\max}$  для ASCII-кода составляет 7 бит, что заметно превышает аналогичный показатель равномерного кода Хартли (5.64 бита) и обусловлено более крупным алфавитом. Для представления символов ASCII требуется 8 бит, тогда как равномерному коду Хартли достаточно всего 6. Абсолютная избыточность ASCII достигает 2.49 бита, в то время как у кода Хартли этот показатель равен лишь 1.14 бита. Относительная избыточность также отличается: у ASCII-кода она составляет 36%, что выше, чем 20% у равномерного кода Хартли, что указывает на меньшую эффективность использования битов. Таким образом, равномерный код Хартли оказывается более экономичным и рациональным благодаря снижению избыточности.

## 2 Кодирование дискретных источников информации методом Шеннона-Фано

На рисунке 2 представлена вычисление кода Шеннона-Фано с помощью написанного приложения.

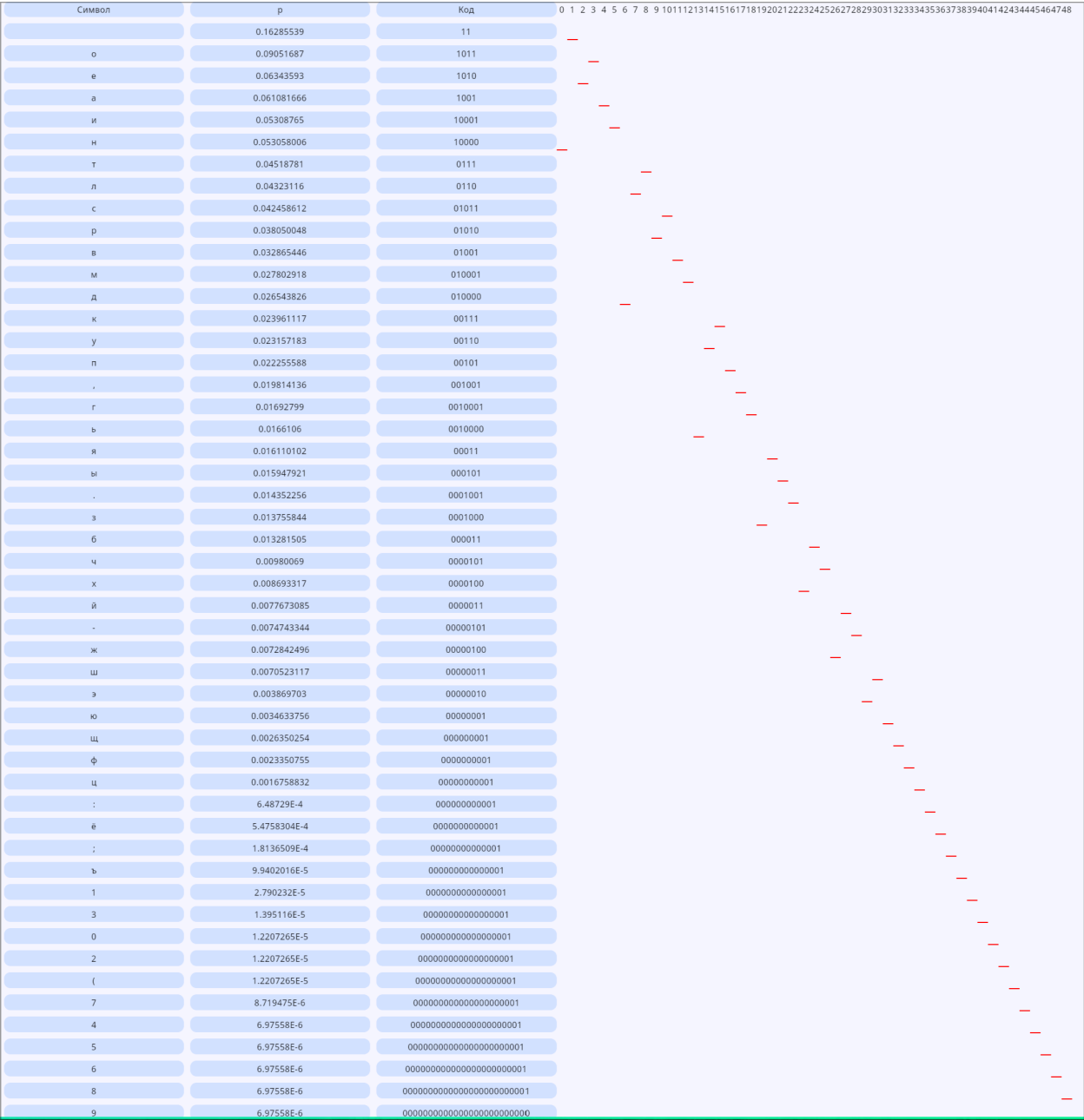


Рисунок 2 — Отображение сгенерированного кода в приложении

Для наглядного отображения данных таблица сгенерированного кода по алгоритму Шеннона-Фано представлена в таблице

Таблица 3 — Код сгенерированный по алгоритму Шеннона-Фано

Символ	р	Код
	0.16285539	11



Символ	р	Код
о	0.09051687	1011
е	0.06343593	1010
а	0.061081666	1001
и	0.05308765	10001
н	0.053058006	10000
т	0.04518781	0111
л	0.04323116	0110
с	0.042458612	01011
р	0.038050048	01010
в	0.032865446	01001
м	0.027802918	010001
д	0.026543826	010000
к	0.023961117	00111
у	0.023157183	00110
п	0.022255588	00101
,	0.019814136	001001
г	0.01692799	0010001
ь	0.0166106	0010000
я	0.016110102	00011
ы	0.015947921	000101
.	0.014352256	0001001
з	0.013755844	0001000
б	0.013281505	000011
ч	0.00980069	0000101
х	0.008693317	0000100
й	0.0077673085	0000011
-	0.0074743344	00000101
ж	0.0072842496	00000100
ш	0.0070523117	00000011
э	0.003869703	00000010
ю	0.0034633756	00000001
щ	0.0026350254	000000001
ф	0.0023350755	0000000001
ц	0.0016758832	00000000001
:	6.48729E-4	000000000001



Получив сообщение подобного вида, необходимо его декодировать, чтобы прочесть. Считаем, что получатель имеет таблицу кодировки символов, идентичную с отправителем. Декодирование производится с наименьшей длины кодового слова, — получается таблица 4 с итоговым результатом, аналогичным закодированному

Таблица 4 — Декодирование кода

Шаг	Комбинация	Кол-во	Символ
1	00	37	-
2	001	6	-
3	0010	4	-
4	00101	1	п
5	10	5	-
6	100	3	-
7	1001	1	а
8	10	5	-
9	100	3	-
10	1000	2	-
11	10000	1	н
12	00	37	-
13	001	6	-
14	0011	2	-
15	00111	1	к
16	10	5	-
17	101	2	-
18	1011	1	о
19	01	7	-
20	010	5	-
21	0100	3	-
22	01001	1	в
23	11	1	
24	01	7	-
25	010	5	-
26	0100	3	-
27	01000	2	-
28	010001	1	м

Шаг	Комбинация	Кол-во	Символ
29	00	37	-
30	000	31	-
31	0000	27	-
32	00000	24	-
33	000000	21	-
34	0000000	19	-
35	00000000	18	-
36	000000000	17	-
37	0000000000	16	-
38	00000000000	15	-
39	000000000000	14	-
40	0000000000000	13	-
41	00000000000000	12	-
42	000000000000000	11	-
43	0000000000000000	10	-
44	00000000000000000	9	-
45	000000000000000000	8	-
46	0000000000000000000	7	-
47	00000000000000000000	6	-
48	000000000000000000000	5	-
49	0000000000000000000001	1	4
50	00	37	-
51	000	31	-
52	0000	27	-
53	00000	24	-
54	000000	21	-
55	0000000	19	-
56	00000000	18	-
57	000000000	17	-
58	0000000000	16	-
59	00000000000	15	-
60	000000000000	14	-
61	0000000000000	13	-
62	00000000000000	12	-
63	000000000000000	11	-

Шаг	Комбинация	Кол-во	Символ
64	000000000000000001	1	1
65	00	37	-
66	000	31	-
67	0000	27	-
68	00000	24	-
69	000000	21	-
70	0000000	19	-
71	00000000	18	-
72	000000000	17	-
73	0000000000	16	-
74	00000000000	15	-
75	000000000000	14	-
76	0000000000000	13	-
77	00000000000000	12	-
78	000000000000000	11	-
79	0000000000000000	10	-
80	00000000000000000	9	-
81	000000000000000000	8	-
82	0000000000000000001	1	2

## 2.2 Вывод

В ходе практического занятия изучался конструктивный подход к созданию кодов для дискретного источника информации, разработанный Клодом Шенноном. Была продемонстрирована работа основной теоремы Шеннона. Созданный код оказался близким к максимально эффективному решению по Шеннону, хотя и не достигал оптимальности. На конкретном примере было показано, что раскодирование переданного сообщения возможно однозначно.

## 3 Кодирование дискретных источников информации методом Д.Хаффмана

На рисунке 4 представлен скриншот из разработанного приложения со сложением вероятностей на каждом шаге.

Рисунок 4 — Посимвольное кодирование методом Хаффмана

По итогу получилась таблица 5.

### Таблица 5 — Таблица кодировки по Хаффману

Символ	p	Код
	0.16285539	001
о	0.09051687	111
е	0.06343593	0110
а	0.061081666	0111
и	0.05308765	1010
н	0.053058006	1011
т	0.04518781	00000
л	0.04323116	00001
с	0.042458612	00010
р	0.038050048	01000
в	0.032865446	01010
м	0.027802918	10010
д	0.026543826	11000
к	0.023961117	11001
у	0.023157183	11010
п	0.022255588	11011

Символ	p	Код
,	0.019814136	000110
г	0.01692799	010010
ь	0.0166106	010011
я	0.016110102	010111
ы	0.015947921	100000
.	0.014352256	100010
з	0.013755844	100110
б	0.013281505	100111
ч	0.00980069	0001110
х	0.008693317	0101100
й	0.0077673085	0101101
-	0.0074743344	1000010
ж	0.0072842496	1000011
ш	0.0070523117	1000110
э	0.003869703	00011111
ю	0.0034633756	10001110
щ	0.0026350254	000111100
ф	0.0023350755	000111101
ц	0.0016758832	100011110
:	6.48729E-4	1000111111
ё	5.4758304E-4	10001111100
;	1.8136509E-4	100011111011
ть	9.9402016E-5	1000111110101
1	2.790232E-5	100011111010001
3	1.395116E-5	1000111110100100
0	1.2207265E-5	1000111110100111
(	1.2207265E-5	10001111101000001
2	1.2207265E-5	10001111101000000
7	8.719475E-6	10001111101000010
6	6.97558E-6	10001111101001101
5	6.97558E-6	10001111101001100
9	6.97558E-6	10001111101001011
8	6.97558E-6	10001111101001010
4	6.97558E-6	10001111101000011

Кодовое дерево представлено на рисунке 5.

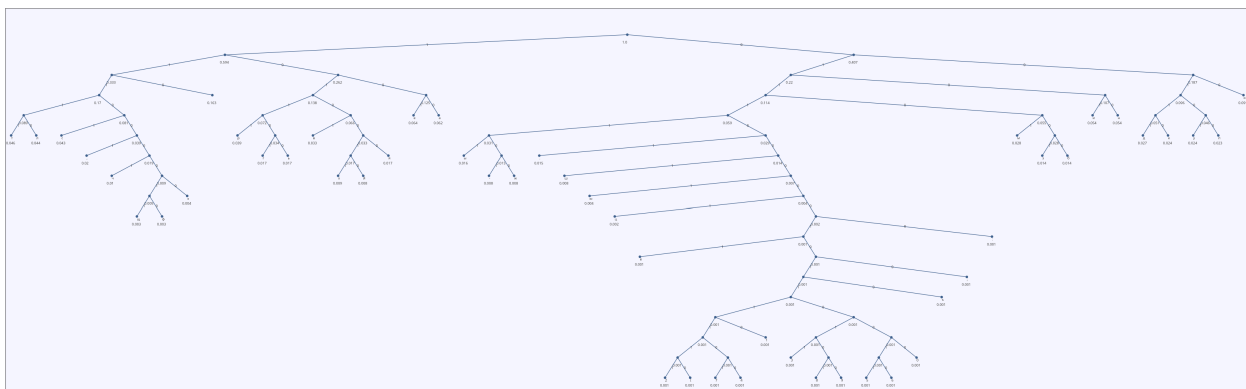


Рисунок 5 — Построенный граф

Изображение графа для подробного просмотра прикреплено к отчёту.

Энтропия источника осталась неизменной и равной 4.51. Рассчитаем среднее количество двоичных разряд. Рассчёт среднего количества двоичных разрядов представлен в формуле (9).

$$n_{\text{ср}} = \sum_{i=1}^{50} p_i a_i \approx 4.54 \quad (9)$$

### 3.1 Пример декодирования сообщения

Закодируем простое сообщение «панков м412». В данном случае воспользуюсь разработанной программой и получу код представленный на рисунке 6.



Рисунок 6 — Сгенерированный код

Получив сообщение подобного вида, необходимо его декодировать, чтобы прочитать. Считаем, что получатель имеет таблицу кодировки символов, идентичную с отправителем. Декодирование производится с наименьшей



длины кодового слова, — получается таблица 6 с итоговым результатом, аналогичным закодированному

Таблица 6 — Декодирование кода

Шаг	Комбинация	Кол-во	Символ
1	110	4	-
2	1101	2	-
3	11011	1	п
4	011	2	-
5	0111	1	а
6	101	2	-
7	1011	1	н
8	110	4	-
9	1100	2	-
10	11001	1	к
11	111	1	о
12	010	7	-
13	0101	4	-
14	01010	1	в
15	001	1	
16	100	25	-
17	1001	3	-
18	10010	1	м
19	100	25	-
20	1000	22	-
21	10001	19	-
22	100011	18	-
23	1000111	17	-
24	10001111	16	-
25	100011111	15	-
26	1000111110	14	-
27	10001111101	13	-
28	100011111010	12	-
29	1000111110100	11	-
30	10001111101000	5	-
31	100011111010000	4	-
32	1000111110100001	2	-

Шаг	Комбинация	Кол-во	Символ
33	10001111101000011	1	4
34	100	25	-
35	1000	22	-
36	10001	19	-
37	100011	18	-
38	1000111	17	-
39	10001111	16	-
40	100011111	15	-
41	1000111110	14	-
42	10001111101	13	-
43	100011111010	12	-
44	1000111110100	11	-
45	10001111101000	5	-
46	100011111010001	1	1
47	100	25	-
48	1000	22	-
49	10001	19	-
50	100011	18	-
51	1000111	17	-
52	10001111	16	-
53	100011111	15	-
54	1000111110	14	-
55	10001111101	13	-
56	100011111010	12	-
57	1000111110100	11	-
58	10001111101000	5	-
59	100011111010000	4	-
60	1000111110100000	2	-
61	10001111101000000	1	2

### 3.2 Вывод

На практике был изучен и применён метод кодирования дискретных источников информации, разработанный Д. Хаффманом. В отличие от рассмотренного ранее метода, подверженного неоднозначности, алгоритм Хаффмана обеспечивает однозначное кодирование. Этот подход позволяет создавать код с

минимально возможным средним количеством двоичных символов для заданного распределения вероятностей.

В процессе построения кодового дерева установлено, что кодовые слова присваиваются исключительно конечным вершинам дерева. Благодаря этому достигается важное свойство: ни одно кодовое слово не является началом другого, что делает возможным точное разделение последовательности на отдельные элементы.

На примере было подтверждено, что декодирование сообщений, закодированных методом Хаффмана, является однозначным.

## **ЗАКЛЮЧЕНИЕ**

В ходе выполнения лабораторной работы №7 были приобретены практические навыки по численному определению количества информации в сообщении, а также освоены методы построения кодов дискретного источника информации с использованием подходов Шеннона-Фано и Хаффмана. Результаты работы подтвердили теоретическую однозначность декодирования сообщений, что продемонстрировано на конкретных примерах. Последовательное выполнение всех трёх частей работы позволило глубже понять принципы кодирования и декодирования информации, а также закрепить полученные знания на практике.