

КАФЕДРА №

ОТЧЕТ
ЗАЩИЩЕН С ОЦЕНКОЙ
ПРЕПОДАВАТЕЛЬ

должность, уч. степень, звание

подпись, дата

инициалы, фамилия

ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ

Лабораторная работа №7 \ 18 Вариант

Оценка количества информации в сообщении и эффективное кодирование.

РАБОТУ ВЫПОЛНИЛ

СТУДЕНТ ГР. №

подпись, дата

инициалы, фамилия

Санкт-Петербург 2021

ЧАСТЬ 1

Определение количества информации, содержащейся в сообщении.

Цель работы:

Получение практических навыков численного определения количества информации, содержащегося в сообщении. Освоение методов построения кодов дискретного источника информации используя конструктивный метод, предложенный К.Шенноном и Н.Фано, и метод Хаффмана. На примере показать однозначность раскодирования имеющегося сообщения.

Материалы:

А.С. Пушкин «Евгений Онегин» 1 и 2 главы.

Заполненная таблица №1 для 50-ти символов.

№ n/n	Символ	Код символа	Число вхождений символа в текст	Вероятность вхождения символа (p _i)	li	lcp
1	0	48	27	0,000809401	10,27085767	0,008313243
2	1	49	90	0,002698003	8,533892077	0,031337713
3	2	50	35	0,001049224	9,896462156	0,041721315
4	3	51	13	0,000389712	11,32530546	0,046134918
5	4	52	26	0,000779423	10,32530546	0,054182701
6	5	53	59	0,001768691	9,143102124	0,070354025
7	6	54	49	0,001468913	9,411035329	0,084178017
8	7	55	99	0,002967804	8,396388553	0,109096851
9	8	56	43	0,001289046	9,599480419	0,121471024
10	9	57	12	0,000359734	11,44078267	0,12558666
11	.	46	252	0,00755441	7,04846525	0,178833655
12	,	44	781	0,023412675	5,416566435	0,305649962
13	:	58	67	0,002008514	8,959655983	0,323645554
14	;	59	158	0,004736495	7,721964425	0,3602206
15	-	45	18	0,000539601	10,85582017	0,366078408
16		32	4266	0,127885365	2,967076923	0,745524123
17	(40	9	0,0002698	11,85582017	0,748722827
18	а	224	2019	0,060525211	4,046319978	0,993627199
19	б	225	504	0,015108819	6,04846525	1,085012368
20	в	226	1253	0,037562204	4,734574474	1,26285342
21	г	227	526	0,015768331	5,986826184	1,35725568
22	д	228	993	0,029767972	5,070095266	1,508182132
23	е	229	2331	0,06987829	3,839011884	1,776445718
24	ё	184	8	0,000239823	12,02574517	1,779329763
25	ж	230	335	0,010042568	6,637727888	1,8459896
26	з	231	475	0,014239463	6,13396147	1,933333916
27	и	232	1775	0,053210624	4,232141864	2,158528826
28	й	233	557	0,016697644	5,904211656	2,257115249
29	к	234	784	0,023502608	5,411035329	2,384288692
30	л	235	1428	0,042808322	4,545964909	2,57889382
31	м	236	947	0,028388992	5,138524558	2,724771354
32	н	237	2047	0,06136459	4,026449786	2,971852795
33	о	238	2757	0,08264884	3,596861621	3,269129235
34	п	239	733	0,021973739	5,508075785	3,390162257
35	р	240	1225	0,036722825	4,767179139	3,565226543

36	с	241	1355	0,040619941	4,621668037	3,752958427
37	т	242	1428	0,042808322	4,545964909	3,947563556
38	у	243	836	0,025061455	5,318386041	4,080850046
39	ф	244	34	0,001019246	9,938282332	4,090979598
40	х	245	300	0,008993345	6,796926483	4,152106702
41	ц	246	109	0,003267582	8,257560848	4,179088959
42	ч	247	395	0,011841237	6,40003633	4,254873309
43	ш	248	211	0,006325319	7,304645985	4,301077527
44	щ	249	50	0,001498891	9,381888983	4,315139955
45	ъ	250	6	0,000179867	12,44078267	4,31737764
46	ы	251	685	0,020534804	5,605784995	4,432491337
47	ь	252	575	0,017237244	5,858327027	4,533472752
48	э	253	32	0,00095929	10,02574517	4,543090351
49	ю	254	181	0,005425985	7,525899286	4,583925765
50	я	255	460	0,013789796	6,180255122	4,66915022

всего символов(K)	Полная вероятность (P)	Энтропия источника
33358	1	4,66915022

Заполненная таблица №2.

	Алфавит	Неопределённость	Разрядность кода	Абсолютная избыточность	Относительная избыточность
Стандартная кодовая таблица ASCII	255	7,994353437	8	-	-
Мера Хартли	50	5,64385619	6	0,97470597	0,172702127

Использованные формулы с определением переменных.

$I_{cp} = \sum p_i (-\log_2 p_i) = H$ — (энтропия по Шеннону, где $p_i = k_i/K$ вероятность вхождения символа в текст; k_i - количество появления символа в тексте, K -количество символов в тексте).

$H_1 = \log_2 N$ — (Исходная неопределенность по Хартли).

$H_2 = \log_2 N^*$ — (Остаточная неопределенность по Хартли, где N^* — число возможных значений принятого слова после получения сообщения). Энтропия по Хартли оценивается как мера неопределенности, учитывающая вероятность появления и информативность того или иного сообщения.

$I_i = -\log_2 p_i$ — (I_i — количество вносимой символом информации; p_i -вероятность вхождения символа).

$D = (H_{max} - H) / H_{max}$ — (H_{max} — максимальное значение энтропии; D — относительная избыточность).

$D_{abc} = H_{max} - H$ — (D_{abc} — абсолютная избыточность).

Выводы.

Были получены практические навыки численного определения количества информации, содержащегося в сообщении. Произведя сравнительный анализ таблицы 2, мы видим, что абсолютная избыточность и относительная избыточность для стандартной кодовой таблицы ASCII и Меры Хартли имеют разные значения, так как они связаны с неопределенностью. Отсюда следует, что лучше использовать Меру Хартли.

ЧАСТЬ 2

Кодирование дискретных источников информации методом Шеннона-Фано.

Использованные формулы с определением переменных.

Энтропия набора символов в рассматриваемом случае определяется как

$$H = -\sum_{i=1}^8 p(a_i) \log_2 p(a_i) \approx 4,722535$$

H –энтропия, $p(a_i)$ -вероятность символа.

Среднее количество двоичных разрядов, используемых при кодировании символов по алгоритму Шеннона-Фано:

$$I_{cp} = \sum_{i=1}^A p(a_i) n(a_i),$$

A — размер (или объем) алфавита, используемого для передачи сообщения; $n(a_i)$ — число двоичных разрядов в кодовой комбинации, соответствующей символу a_i .

Разбиение букв.

Символ	Вероятность вхождения символа (p _i)	Кодовая комбинация	1	2	3	4	5	6	7	8	9	##	##	##	Число двоичных разрядов
	0,127885365	111	1	1											3
о	0,08264884	1101	1	0	1										4
е	0,06987829	1100	1	0	0										4
н	0,06136459	1011	1	1	1										4
а	0,060525211	1010	1	1	0										4
и	0,053210624	1001	1	1	0	1									4
л	0,042808322	1000	1	1	0	0									4
т	0,042808322	0111	1	1	0	0									4
с	0,040619941	01101	1	1	0	1									5
в	0,037562204	01100	1	1	0	0									5
р	0,036722825	01011	1	1	0	1									5
д	0,029767972	01010	1	1	0	0									5
м	0,028388992	01001	1	1	0	1									5
у	0,025061455	01000	1	1	0	0									5
к	0,023502608	00111	1	1	0	1									5
г	0,023412675	001101	1	1	0	1	1								6
п	0,021973739	001100	1	1	0	1	0								6
ы	0,020534804	00101	1	1	0	1	0								5
ь	0,017237244	001001	1	1	0	1	0	1							6
й	0,016697644	001000	1	1	0	1	0	0							6
г	0,015768331	000111	1	1	0	1	0	1							6
б	0,015108819	000110	1	1	0	1	0	0							6
з	0,014239463	000101	1	1	0	1	0	1							6
я	0,013789796	0001001	1	1	0	1	0	1	1						7
ч	0,011841237	0001000	1	1	0	1	0	1	0						7
ж	0,010042568	0000111	1	1	0	1	0	1	0						7
х	0,008993345	0000110	1	1	0	1	0	1	0						7
.	0,00755441	0000101	1	1	0	1	0	1	0						7
ш	0,006325319	0000100	1	1	0	1	0	1	0						7
ю	0,005425985	00000111	1	1	0	1	0	1	0	1					8
;	0,004736495	00000110	1	1	0	1	0	1	0	0					8
ц	0,003267582	00000101	1	1	0	1	0	1	0	0					8
7	0,002967804	00000100	1	1	0	1	0	1	0	0					8
1	0,002698003	000000111	1	1	0	1	0	1	0	0	1				9
:	0,002008514	000000110	1	1	0	1	0	1	0	0	0				9
5	0,001768691	000000101	1	1	0	1	0	1	0	0	0				9
щ	0,001498891	000000100	1	1	0	1	0	1	0	0	0				9
6	0,001468913	0000000111	1	1	0	1	0	1	0	0	1				10
8	0,001289046	0000000110	1	1	0	1	0	1	0	0	0				10
2	0,001049224	0000000101	1	1	0	1	0	1	0	0	0				10
ф	0,001019246	0000000100	1	1	0	1	0	1	0	0	0				10
э	0,00095929	0000000011	1	1	0	1	0	1	0	0	0				10
0	0,000809401	00000000101	1	1	0	1	0	1	0	0	1				11
4	0,000779423	00000000100	1	1	0	1	0	1	0	0	0				11
-	0,000539601	00000000011	1	1	0	1	0	1	0	0	0				11
3	0,000389712	00000000010	1	1	0	1	0	1	0	0	0				11
9	0,000359734	000000000011	1	1	0	1	0	1	0	0	0	1			12
(0,0002698	000000000010	1	1	0	1	0	1	0	0	0	0			12
в	0,000239823	000000000001	1	1	0	1	0	1	0	0	0	0			12
ь	0,000179867	000000000000	1	1	0	1	0	1	0	0	0	0			12

Символ	Кодовая комбинация	Число двоичных разрядов	Вероятность вхождения символа (p _i)	Вероятность дл. Строки	Ср. кол-во символов(пср)	"-log(p _i) по осн 2"	Вычисления	Энтропия H
	111	3	0,127885365	0,38365609		2,967076923	0,37944571	
о	1101	4	0,08264884	0,33059536		3,596861621	0,29727644	
е	1100	4	0,06987829	0,27951316		3,839011884	0,26826359	
н	1011	4	0,06136459	0,24545836		4,026449786	0,24708144	
а	1010	4	0,060525211	0,24210085		4,046319978	0,24490437	
и	1001	4	0,053210624	0,2128425		4,232141864	0,22519491	
л	1000	4	0,042808322	0,17123329		4,545964909	0,19460513	
т	0111	4	0,042808322	0,17123329		4,545964909	0,19460513	
с	01101	5	0,040619941	0,20309971		4,621668037	0,18773188	
в	01100	5	0,037562204	0,18781102		4,734574474	0,17784105	
р	01011	5	0,036722825	0,18361413		4,767179139	0,17506429	
д	01010	5	0,029767972	0,14883986		5,070095266	0,15092645	
м	01001	5	0,028388992	0,14194496		5,138524558	0,14587753	
у	01000	5	0,025061455	0,12530727		5,318386041	0,13328649	
к	00111	5	0,023502608	0,11751304		5,411035329	0,12717344	
г	001101	6	0,023412675	0,14047605		5,416566435	0,12681631	
п	001100	6	0,021973739	0,13184244		5,508075785	0,12103302	
ы	00101	5	0,020534804	0,10267402		5,605784995	0,1151137	
ь	001001	6	0,017237244	0,10342347		5,858327027	0,10098141	
й	001000	6	0,016697644	0,10018586		5,904211656	0,09858642	
г	000111	6	0,015768331	0,09460999		5,986826184	0,09440226	
б	000110	6	0,015108819	0,09065292		6,04846525	0,09138517	
з	000101	6	0,014239463	0,08543678		6,13396147	0,08734432	
я	0001001	7	0,013789796	0,09652857		6,180255122	0,08522445	
ч	0001000	7	0,011841237	0,08288866	4,704988309	6,40003633	0,07578435	
ж	0000111	7	0,010042568	0,07029798		6,637727888	0,06665984	4,66915022
х	0000110	7	0,008993345	0,06295341		6,796926483	0,0611271	
.	0000101	7	0,00755441	0,05288087		7,04846525	0,05324699	
ш	0000100	7	0,006325319	0,04427723		7,304645985	0,04620422	
ю	00000111	8	0,005425985	0,04340788		7,525899286	0,04083541	
;	00000110	8	0,004736495	0,03789196		7,721964425	0,03657505	
ц	00000101	8	0,003267582	0,02614066		8,257560848	0,02698226	
7	00000100	8	0,002967804	0,02374243		8,396388553	0,02491883	
1	000000111	9	0,002698003	0,02428203		8,533892077	0,02302447	
:	000000110	9	0,002008514	0,01807662		8,959655983	0,01799559	
5	000000101	9	0,001768691	0,01591822		9,143102124	0,01617132	
щ	000000100	9	0,001498891	0,01349002		9,381888983	0,01406243	
6	0000000111	10	0,001468913	0,01468913		9,411035329	0,01382399	
8	0000000110	10	0,001289046	0,01289046		9,599480419	0,01237417	
2	0000000101	10	0,001049224	0,01049224		9,896462156	0,0103836	
ф	0000000100	10	0,001019246	0,01019246		9,938282332	0,01012955	
э	0000000011	10	0,00095929	0,0095929		10,02574517	0,00961676	
0	00000000101	11	0,000809401	0,00890341		10,27085767	0,00831324	
4	00000000100	11	0,000779423	0,00857366		10,32530546	0,00804778	
-	00000000011	11	0,000539601	0,00593561		10,85582017	0,00585781	
3	00000000010	11	0,000389712	0,00428683		11,32530546	0,0044136	
9	000000000011	12	0,000359734	0,00431681		11,44078267	0,00411564	
(000000000010	12	0,0002698	0,0032376		11,85582017	0,0031987	
в	000000000001	12	0,000239823	0,00287787		12,02574517	0,00288404	
ь	000000000000	12	0,000179867	0,0021584		12,44078267	0,00223768	

Составленное сообщение, содержащее не менее 10 символов и кодовая строка:

Сообщение:	информатика .
	1001.1011.0000000100.1101.01011.01001.1010.0111.1001.00111.1010.111.0000101.

Декодирование			
Шаг	Комбинация	Кол-во символов	Символ
1	100	2	-
2	1001	1	и
3	101	2	-
4	1011	1	н
5	000	30	-
6	0000	25	-
7	00000	21	-
8	000000	17	-
9	0000000	13	-
10	00000001	4	-
11	000000010	2	-
12	0000000100	1	ф
13	110	2	-
14	1101	1	о
15	010	4	-
16	0101	2	-
17	01011	1	р
18	010	4	-
19	0100	2	-
20	01001	1	м
21	101	2	-
22	1010	1	а
23	011	3	-
24	0111	1	т
25	100	2	-
26	1001	1	и
27	001	6	-
28	0011	3	-
29	00111	1	к
30	101	2	-
31	1010	1	а
32	111	1	и и
33	000	30	-
34	0000	25	-
35	00001	4	-
36	000010	2	-
37	0000101	1	-

Вывод: Я освоил метод построения кода дискретного источника информации, используя конструктивный метод, предложенный К. Шенноном и Н.Фано.

Очевидно, передача сообщения с помощью такого кода будет более эффективной, т.к. будет требовать меньшего количества бинарных сигналов при кодировании. Можно говорить о том, что при таком кодировании мы не передаем избыточную информацию, которая была бы в восьмибитовом кодировании. Кодирование и декодирование однозначно.

ЧАСТЬ 3

Кодирование дискретных источников информации методом Д.Хаффмана.

49 шагов суммирования вероятностей:

[illegible][illegible]

[illegible]

Рисунок кодового дерева, с полученными значениями и подписанными символами.

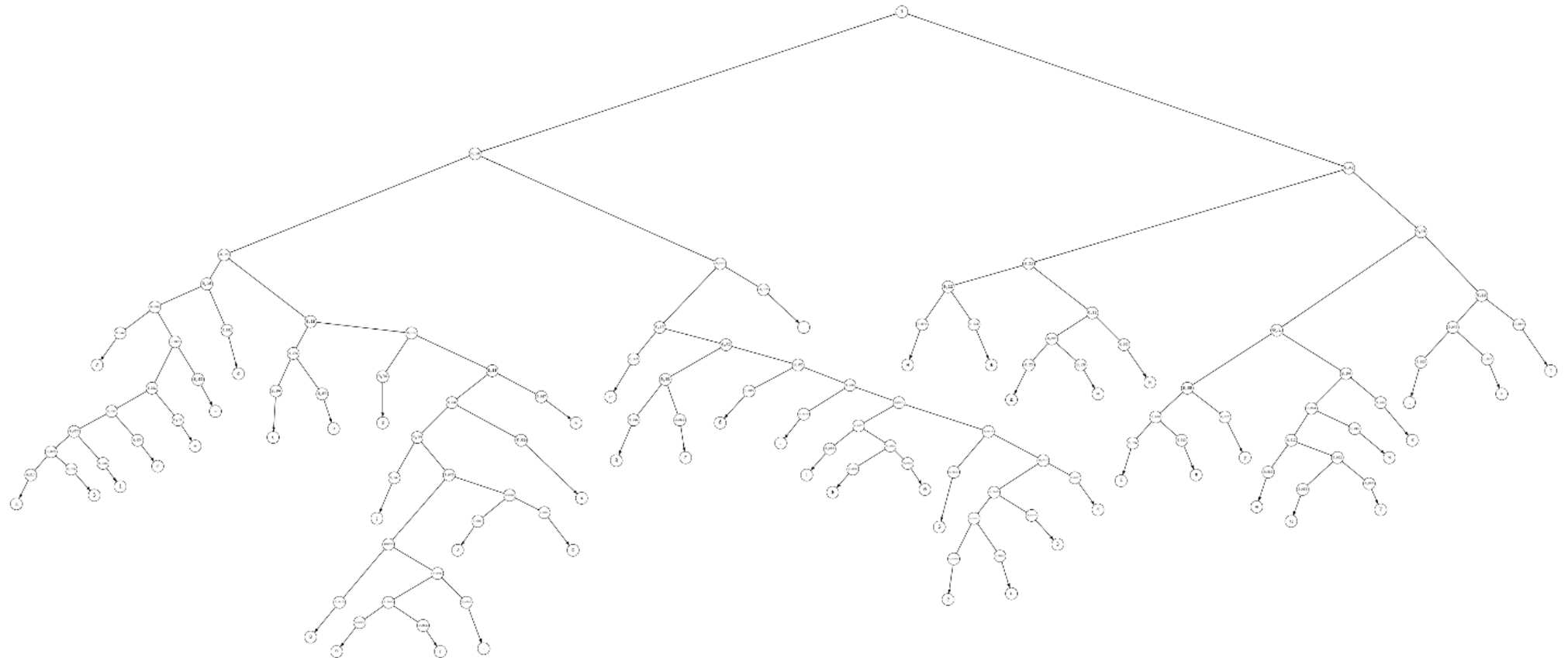


Таблица полученных кодов.

Символ	Код
	100
о	1110
е	1011
н	0111
а	0110
и	0100
л	11111
т	0000
с	11011
в	11010
р	11001
д	01011
м	01010
у	00110
к	00100
,	00011
п	00010
ы	111100
ь	110000
й	101011
г	101010
б	001111
з	1111011110
я	001110
ч	001010
ж	1111010
х	1100010
.	1010001
ш	00101111
ю	11110110
;	11000111
ц	00101101
7	00101100
1	111101110
:	101000011

5	101000001
щ	1111011111
6	101001
8	1100011011
2	1100011001
ф	1100011000
э	1010000101
0	1010000100
4	1010000000
-	11000110100
3	10100000010
9	110001101011
(110001101010
ё	101000000111
ь	101000000110

Составленное сообщение, содержащее не менее 10 символов и кодовая строка.

Сообщение:	информатика .
	0100.0111.1100011000.1110.11001.01010.0110.0000.0100.00100.011.100.1010001.

Описание декодирования данного сообщения.

Декодирование			
Шаг	Комбинация	Кол-во символов	Символ
1	010	3	-
2	0100	1	и
3	011	2	-
4	0111	1	н
5	110	12	-
6	1100	10	-
7	11000	9	-
8	110001	8	-
9	1100011	7	-
10	11000110	6	-
11	110001100	2	-
12	1100011000	1	ф
13	111	8	-
14	1110	1	о
15	110	12	-
16	1100	10	-
17	11001	1	р
18	010	3	-
19	0101	2	-
20	01010	1	м
21	011	2	-
22	0110	1	а
23	000	3	-
24	0000	1	т
25	010	3	-
26	0100	1	и
27	001	7	-
28	0010	2	-
29	00100	1	к
30	011	2	-
31	0110	1	а
32	100	1	" "
33	101	13	-
34	1010	12	-
35	10100	11	-
36	101000	10	-
37	1010001	1	.

Вывод: Мы освоили метод построения кодов дискретного источника информации используя методику Д.Хаффмана. На примере показали однозначность раскодирования имеющегося сообщения