

Практическая работа №8.1

Тема: кодирование и сжатие данных методами без потерь.

Цель: получение практических навыков и знаний по выполнению сжатия данных рассматриваемыми методами

Задание 1. Исследование алгоритмов сжатия на примерах.

- 1) Выполнить каждую задачу варианта, представив алгоритм решения в виде таблицы и указав результат сжатия. Примеры оформления решения представлены в Приложении1 этого документа.
- 2) Описать процесс восстановления сжатого текста.
- 3) Сформировать отчет, включив задание, вариант задания, результаты выполнения задания варианта.

Алгоритм 1. Применение алгоритма группового сжатия текста RLE.

Сжать текст, используя метод RLE (run length encoding/кодирование длин серий/групповое кодирование).

Требования к выполнению заданию

- 1) Описать процесс сжатия алгоритмом RLE.
- 2) Придумать текст, в котором есть длинные (в разумных пределах) серии из повторяющихся символов. Выполнить сжатие текста. Рассчитать коэффициент сжатия.
- 3) Придумать текст, в котором много неповторяющихся символов и между ними могут быть серии. Выполнить групповое сжатие, показать коэффициент сжатия. Применить алгоритм разделения текста при групповом кодировании, позволяющий повысить эффективность сжатия этого текста. Рассчитать коэффициент сжатия после применения алгоритма.

Алгоритм 2. Исследование алгоритмов группового сжатия (методы Лемпеля –Зива: LZ77, LZ78) на примерах.

Тексты для сжатия по вариантам в таблице 1 в столбце 1 (для LZ77) и в столбце 2 (для LZ78).

Требования к выполнению заданию

- 1) Выполнить каждую задачу варианта задания, представив алгоритм решения в виде таблицы и указав результат сжатия.
- 2) Описать процесс восстановления сжатого текста.

3) Сформировать отчет, включив задание, вариант задания, результаты выполнения задания варианта (таблица).

Алгоритм 3,4. Разработать на псевдокоде или описать словесно алгоритмы сжатия и восстановления текста методами Шеннона-Фано и Хаффмана. Представить процесс выполнения алгоритма для задачи варианта. Все результаты представить в отчете.

Требования к выполнению заданию

1. Сформировать отчет по разработке каждого алгоритма в соответствии с требованиями.

1.1. По методу Шеннона-Фано.

- 1) Данными для выполнения задания является текст, представленный в таблице 1 в столбце 3.
- 2) Привести постановку задачи, описать алгоритм формирования префиксного дерева и алгоритм кодирования, декодирования.
- 3) Представить таблицу формирования кода для текста варианта задания.
- 4) Изобразить префиксное дерево формирования кода.
- 5) Рассчитать коэффициент сжатия.

1.2. По методу Хаффмана.

- 1) Данные для выполнения задания: ваша фамилия имя отчество.
- 2) Привести постановку задачи, описать алгоритм формирования префиксного дерева и алгоритм кодирования, декодирования.
- 3) Построить таблицу частот встречаемости символов в исходной строке для чего сформировать алфавит исходной строки и посчитать количество вхождений (частот) символов и их вероятности появления.
- 4) Изобразить префиксное дерево Хаффмана.
- 5) Упорядочить построенное дерево слева-направо (при необходимости) и изобразить его.
- 6) Провести кодирование исходной строки по аналогии с примером:
- 7) Рассчитать коэффициенты сжатия относительно кодировки ASCII и относительно равномерного кода.
- 8) Рассчитать среднюю длину полученного кода и его дисперсию.
- 9) По результатам выполненной работы сделать выводы и сформировать отчет. Отобразить результаты выполнения всех требований (с 1 по 8), предъявленных в задании и оформить разработку программы: постановка, подход к решению, код, результаты тестирования.

Задание 2. Реализовать и отладить программу.

1. Разработать алгоритм и реализовать программу сжатия текста алгоритмом Шеннона – Фано. Разработать алгоритм и программу восстановления сжатого текста. Выполнить тестирование программы на текстовом файле. Определить процент сжатия.
2. Применить алгоритм Хаффмана для архивации данных текстового файла. Выполнить практическую оценку сложности алгоритма Хаффмана. Провести архивацию этого же файла любым архиватором. Сравнить коэффициенты сжатия разработанного алгоритма и архиватора.

Таблица 1. Варианты к заданиям 1

Вариант	Сжатие данных по методу Лемпеля–Зива LZ77 Используя двухсимвольный алфавит (0, 1) закодировать следующую фразу:	Закодировать следующую фразу, используя код LZ78	Закодировать фразу методами Шеннона–Фано
	1	2	3
1	0001010010101001101	кукуркукурекурекун	Ана, дэус, рики, паки, Дормы кормы консунтаки, Дэус дэус канадэус – бац!
2	0100100010010000101	упупапекапекаууп	One, two, Freddy's coning for you Three, four, aetter lock your door Five, six, graa a crucifix Seven, eight, gonna stayup late.
3	0100101010010000101	лорлоралоранранлоран	Эне-бене, рики-таки, Буль-буль-буль, Караки-шмаки Эус- деус-краснодеус бац

4	0100001000000100001	пропронепронепрнепрона с	Кони-кони, коникони, Мы сидели на балконе, Чай пили, воду пили, По- турецки говорили.
5	10100010010101000101 1	какатанекатанекатата	Прибавь к ослиной голове Еще одну, получишь две. Но сколько б ни было ослов, Они и двух не свяжут слов.
6	000101110110100111	менменаменаменатеп	По-турецки говорили. Чяби, чяряби Чяряби, чяби-чяби. Мы набралив рот воды.
7	11010101100110000100 1	долделдолдилделдил	Тише, мыши, кот на крыше, А котята ещё выше. Кот пошёл за молоком,А котята кувырком.
8	01011011011010001000 1	sarsalsarsanlasanl 33	Мой котёнок очень странный, Он не хочет есть сметану,К молоку не прикасался И от рыбки отказался.
9	00010010110010001000 1	kloklonkolonklonkl	Эни-бени рити-Фати. Дорба, дорба сентибрати. Дэл. Дэл. Кошка. Дэл. Фати!
10	1110100110110001101	tertrektekertektrek	Самолёт-вертолёт! Посади меня в полёт!А в полёте пусто – Выросла капуста.

11	10101001101100111010	aigaoneaigaoreaigao	Кот пошёл за молоком, А котят кувырком. Кот пришёл без молока, А котят ха-ха-ха.
12	0001001010101001101	connerconneconnerce	Цветом мой зайчишка – белый, А ещё, он очень смелый! Не боится он лисицы, Льва он тоже не боится.
13	01011011001010101011	weaweraweaerweaerwea	Эне, бене, лики, паки, Цуль, буль-буль, Калики-цваки, Эус- беус, кликмадеус, бокс.
14	0010100110010000001	porpoterpoterporter	Ана-дэус-рики-паки, Дормы-кормыконсту- таки, Энус-дэус-кана- дэусБАЦ!
15	10110111100110001101	nantopnentoantonen	Раз, два – упала гора; три, четыре – прицепило; пять, шесть – бьют шерсть; семь, восемь – сено косим.
16	0100101010010000101	roporopoterpoterter	Зуба зуба, зуба зуба, Зуба дони дони мэ, А шарли буба раз два три, А ми раз два три замри.

17	0001000010101001101	weaweraweawerweawerwea	Плыл по морю чемодан, В чемодане был диван, На диване ехал слон. Кто не верит – выйди вон!
18	1110100110111001101	sionsinossionsinos	Дрынцы- брынцыбубен- цы, Раз- звонилисьудальцы, Диги-диги-диги-дон, Выхо-ди-скорее-вон!
19	0001000010101001101	conconconconacon	Перводан, другодан,На колоде барабан; Свистель, коростель, Пятерка, шестерка, утюг.
20	0100101010010000101	nantopnentoponantonen	Эни бэни рики паки Турбаурбасентибряки. Может – выйдет, может – нет, В общем – полный Интернет
21	1010011011000110111	terhrekhekerhekhrek	Матросская шапка, Верёвка в руке, Тяну я кораблик По быстрой реке. И скачут лягушкиЗа мной по пятам, И просят меня: — Прокати, капитан!
22	10100110110011101001	aigaoneaigaoreaigao	Спать пора! Уснул бычок, Лёг в коробку на бочок. Сонный мишка лёг в кровать, Только слон не хочет спать. Головой кивает слон, Он слонихе шлёт поклон.

23	1001010101001101000	connerconneconnerce	Я свою сестренку Лиду Никому не дам в обиду! Я живу с ней очень дружно, Очень я ее люблю. А когда мнебудет нужно, Я и сам ее побью.
24	11011001010101011010	werwerrwererwererwer	Как большой, сидит Андрюшка На ковре перед крыльцом. У него в руках игрушка — Погремушка с бубенцом. Мальчик смотрит — что за чудо? Мальчик очень удивлён, Не поймёт он:ну откуда Раздаётся этот звон?
25	0100110010000001001	porpoherpoherporher	Нет, напрасно мы решили Прокатить кота в машине: Кот кататься не привык —Опрокинул грузовик.
26	10111100110001101101	nanhornenhorokanhoken	Наша Таня громко плачет: Уронила в речку мячик. — Тише,Танечка, не плачь: Не утонет в речке мяч.
27	0101010010000101010	rorororoherroroherher	У меня живёт козлёнок, Я сама его пасу. Я козлёнка в садзелёный Рано утром отнесу. Он заблудитсяв саду — Я в траве его найду.

28	1000010101001101000	сеасегасеаерсеаерсеа	Плыл по морю чемодан, В чемодане был диван, На диване ехал слон. Кто не верит – выйди вон!
29	1010011011100110111	суонсуноссунос	Воробей по лужице Прыгает и кружится. Перышки взъерошил он, Хвостик распушил. Погода хорошая! Чил- чив-чил!
30	0100001010100110100	аокаонаокаонааок	Пять зелёных лягушат В воду броситься спешат — Испугались цапли! А меня они смешат: Я же этой цапли Не боюсь ни капли!

Приложение 1. Примеры оформления выполнения задания 2

1. Для метода Лемпеля –Зива LZ77 для сжатия двоичного кода

Исходный текст	00000001111111111110 00000000011011110
LZ-код	0.00.100.001.011.1011.1101. 1010.0110.10010.10001.10110.
R	2 3 4
Вводимые коды	– 10 11 100 101 110 111 1000 1001 1010 1011 1100

Где LZ – сжатый текст (в данном примере в связи с небольшим размером исходного текста размер текста не уменьшился)

R отмечает шаги кодирования, после которых происходит переход на представление кодов A увеличенным числом разрядов R. Так, на первом шаге вводится код 10 для комбинации 00, и поэтому на следующих двух шагах $R = 2$, после третьего шага $R = 3$, после седьмого шага $R = 4$, т.е. в общем случае $R = K$ после шага $2^{K-1} - 1$.

2. Для метода Лемпеля –Зива LZ78 для сжатия текста

В отличие от LZ77, работающего с уже полученными данными, LZ78 ориентируется на данные, которые только будут получены (LZ78 не использует скользящее окно), он хранит словарь из уже просмотренных фраз.

Пример 1. Дан текст аааааааа. Сжать текст, используя метод LZ78
Словарь

Ссылка на символ	Символы словаря	код
1	а	<0,а>
2	а	<0,а>
3	аа	<1,а>
4	аа	<1,а>
5	ааа	<3,а>

Содержимое словаря	Содержимое считанной строки	Код
а	а	<0,а>

a, a	a	<0,a>
a, a, aa	aa	<1,a>
a, a, aa, aa	aa	<1,a>
a, a, aa, aa, aaa	aaa	<3,a>

Код 0a0a1a1a3a

aaaaaaaaa

Пример 2. Сжать текст saaaaaaaaak. Другая форма таблицы.

Содержимое словаря		Содержимое считанной строки	Код
	1	c	<0,c>
c,	2	a	<0,a>
c,a	3	a	<0,a>
c, a, a	4	aa	<2,a>
c, a, a, aa	5	aaa	<4,a>
c, a, a, aa, aaa	6	aa	<3,a>
c, a, a, aa, aaa, aa	7	aaa	<6,a>
c, a, a, aa, aaa, aa, aaaa		k	<0,k>

Результат сжатия: 0c0a0a2a4a3a5a0k

Приложение 2. Оформление отчета по алгоритмам Шеннона-Фано и Хаффмана.

1. По методу Шеннона-Фано

1) Оформление таблицы метод Шеннона-Фано. Закодирована фраза «Тише, мыши, тише, кот на крыше», используя метод Шеннона-Фано.

Таблица 1

Символ	Кол-во	1-я цифра	2-я цифра	3-я цифра	4-я цифра	5-я цифра	Код	Кол-во бит
пробел	5	0	0	0			000	15
ш	4	0	0	1			001	12
е	3	0	1	0			010	9
,	3	0	1	1			011	9
и	3	1	0	0			100	9

т	3	1	0	1	0		1010	12
ы	2	1	0	1	1		1011	8
к	2	1	1	1	0		1110	8
н	1	1	1	1	1		1111	4
о	1	1	1	0	0	0	11000	5
а	1	1	1	0	0	1	11001	5
м	1	1	1	0	1	0	11010	5
р	1	1	1	0	1	1	11011	5
								106

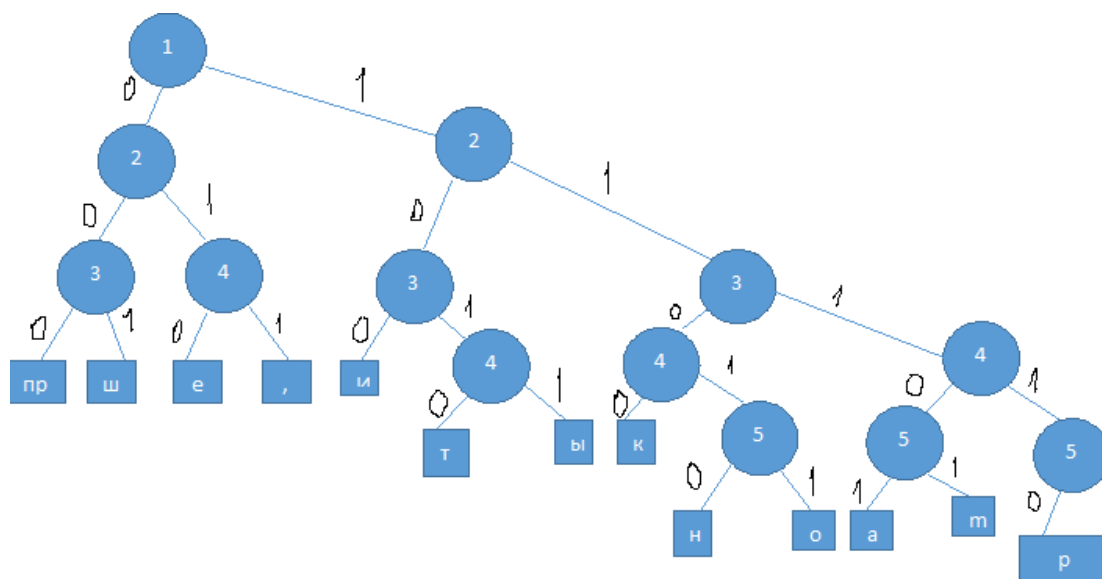
2) Оценка объема и коэффициента сжатия

Объем незакодированной фразы – $30 \cdot 8 \text{ бит} = 240 \text{ бит}$.

Объем закодированной фразы – 106 бит.

Коэффициент сжатия: $106/240=0,44$

3) Дерево префиксного кода Фано



2. По методу Хаффмана

Провести кодирование(сжатие) исходной строки символов «Фамилия Имя Отчество» с использованием алгоритма Хаффмана. Исходная строка символов, таким образом, определяет индивидуальный вариант задания для каждого студента.

Для выполнения задания 3 необходимо выполнить следующие действия:

Исходная строка примера **пупкин василий кириллович**

4. Упорядочить построенное дерево слева-направо (при необходимости).

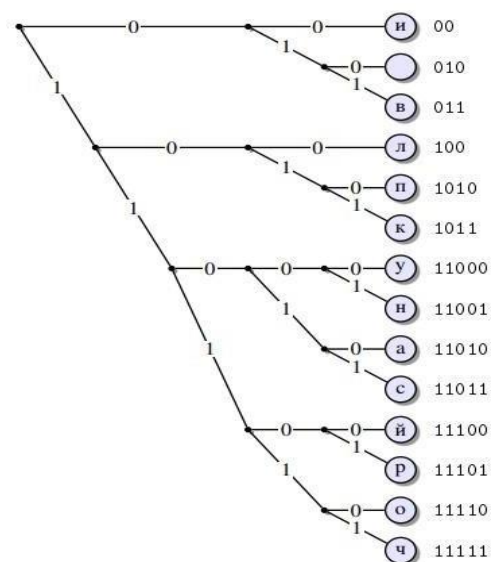


Рисунок. Упорядоченное дерево кодирования Хаффмана

5. Провести кодирование исходной строки по аналогии с примером:

п	у	п	к	и	н	«	»	в	а	с	и	л	и	й
1010	11000	1010	1011	00	11001	010	011	11010	11011	00	100	00	11100	
«	»	к	и	р	и	л	л	о	в	и	ч			
010	1011	00	11101	00	100	100	11110	011	00	11111				

6. Рассчитать коэффициенты сжатия относительно кодировки ASCII и относительно равномерного кода.