**ЗАДАНИЕ 5. КОДИРОВАНИЕ И ДЕКОДИРОВАНИЕ ИНФОРМАЦИИ**

**Кодирование** – это перевод информации с одного языка на другой.

**Декодирование** – обратный переход. Один символ исходного сообщения может заменяться одним или несколькими символами нового кода, а может быть и наоборот – несколько символов исходного сообщения заменяются одним символом в новом коде.

**Кодирование может быть**

|  |  |
| --- | --- |
|  **Равномерное**при равномерном кодировании все символы кодируются кодами равной длины; | **Неравномерное**:при неравномерном кодировании разные символы могут кодироваться кодами разной длины, и это затрудняет однозначное декодирование или делает его невозможным. |

**Условие Фано** означает, что никакое кодовое слово не является началом другого кодового слова. Это обеспечивает возможность однозначной расшифровки закодированных сообщений.

**Обратное условие Фано** также является достаточным условием однозначного декодирования неравномерного кода. В нём требуется, чтобы *никакой код не был окончанием другого (более длинного) кода.*
Для возможности однозначного декодирования достаточно выполнения одного из условий — или прямого, или обратного. Однако существуют варианты неравномерного кодирования, для которых оба условия нарушены, и тем не менее они однозначно декодируются.

**Кодовое дерево (дерево кодирования Хаффмана)** – это двоичное дерево, у которого:

* листья помечены символами, для которых разрабатывается кодировка;
* узлы (в том числе корень) помечены суммой вероятностей появления всех символов, соответствующих листьям поддерева, корнем которого является соответствующий узел.

Метод Хаффмана на входе получает таблицу частот встречаемости символов в исходном тексте. Далее на основании этой таблицы строится дерево кодирования Хаффмана.

Алгоритм построения дерева Хаффмана:

Шаг 1. Символы входного алфавита образуют список свободных узлов. Каждый лист имеет вес, который может быть равен либо вероятности, либо количеству вхождений символа в сжимаемый текст.

Шаг 2. Выбираются два свободных узла дерева с наименьшими весами.

Шаг 3. Создается их родитель с весом, равным их суммарному весу.

Шаг 4. Родитель добавляется в список свободных узлов, а двое его детей удаляются из этого списка.

Шаг 5. Одной дуге, выходящей из родителя, ставится в соответствие бит 1, другой – бит 0.

Шаг 6. Повторяем шаги, начиная со второго, до тех пор, пока в списке свободных узлов не останется только один свободный узел. Он и будет считаться корнем дерева.

Существует два подхода к построению кодового дерева: от корня к листьям и от листьев к корню.

**Пример №1 с решением**

Для передачи по каналу связи сообщения, состоящего только из букв А, Б, В, Г, решили использовать неравномерный по длине код: A=1, Б=01, В=001. Как нужно закодировать букву Г, чтобы длина кода была минимальной, и допускалось однозначное разбиение кодированного сообщения на буквы?

1) 0001

2) 000

3) 11

4) 101

**Решение:**

Это задание удобнее решать с помощью дерева Хаффмана; условие Фано выполняется тогда, когда все выбранные кодовые слова заканчиваются в листьях дерева. По листьям дерева можно однозначно определить, где может находиться буква Г, чтобы длина кода была минимальной, и допускалось однозначное разбиение кодированного сообщения на буквы.

1

0

1

0

1

0

Штриховой линией отмечена «пустая» ветка, на которой можно «прикрепить» лист для кодового слова буквы Г: 2) 000.

**Пример №2 с решением**

По каналу связи передаются сообщения, содержащие только 4 буквы П, О, С, Т; для передачи используется двоичный код, допускающий однозначное декодирование. Для букв Т, О, П используются такие кодовые слова: Т - 111, О - 0, П - 100.

Укажите кратчайшее кодовое слово для буквы С, при котором код будет допускать однозначное декодирование. Если таких кодов несколько, укажите код с наименьшим числовым значением.

**Решение:**

В дереве кода все кодовые слова должны располагаться в листьях дерева, то есть в узлах, которые не имеют потомков. Построим дерево для заданных кодовых слов О – 0, Т – 111 и П – 100:

1

0

1

0

0

0

1

1

Штриховыми линиями отмечены две «пустые» ветви, на которые можно «прикрепить» лист для кодового слова буквы С: 101 или 110; из них минимальное значение имеет код 101.

1

0

1

0

0

0

1

1

**Пример №3 с решением**

Для кодирования некоторой последовательности, состоящей из букв К, Л, М, Н, решили использовать неравномерный двоичный код, удовлетворяющий условию Фано. Для буквы Л использовали кодовое слово 1, для буквы М — кодовое слово 011. Какова наименьшая возможная суммарная длина всех четырёх кодовых слов?

1) 10

2) 9

3) 8

4) 7

**Решение:**

Построим дерево для заданных кодовых слов Л – 1 и М – 011

1

0

1

0

0

1

Штриховыми линиями отмечены две «пустые» ветви, на которые можно «прикрепить» листья для кодовых слов букв К (00) и Н (010)

1

0

1

0

0

1

таким образом, выбрав кодовые слова Л – 1, М – 011, К – 00, Н – 010, получаем суммарную длину кодовых слов 9 символов

Ответ: 2.

**Пример №4 с решением**

Для ко­ди­ро­ва­ния букв Д, X, Р, О, В ре­ши­ли ис­поль­зо­вать дво­ич­ное пред­став­ле­ние чисел 0, 1, 2, 3 и 4 со­от­вет­ствен­но (с со­хра­не­ни­ем од­но­го не­зна­ча­ще­го нуля в слу­чае одноразряд­но­го пред­став­ле­ния). За­ко­ди­руй­те по­сле­до­ва­тель­ность букв ХО­РО­ВОД таким спо­со­бом и ре­зуль­тат за­пи­ши­те вось­ме­рич­ным кодом.

**Решение:**

Сна­ча­ла сле­ду­ет пред­ста­вить дан­ные в усло­вии числа в дво­ич­ном коде:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Д | Х | Р | О | В |
| 0 | 1 | 2 | 3 | 4 |
| 00 | 01 | 10 | 11 | 100 |

Затем за­ко­ди­ро­вать по­сле­до­ва­тель­ность букв: ХО­РО­ВОД — 011110111001100. Те­перь разобьём это пред­став­ле­ние на трой­ки спра­ва на­ле­во и пе­ре­ведём по­лу­чен­ный набор чисел в де­ся­тич­ный код, затем в вось­ме­рич­ный (вось­ме­рич­ное представление сов­па­да­ет с де­ся­тич­ным при раз­би­е­нии трой­ка­ми)

011 110 111 001 100 — 36714.

**Пример №5 с решением**

Для ко­ди­ро­ва­ния букв А, Б, В, Г ре­ши­ли ис­поль­зо­вать двух­раз­ряд­ные по­сле­до­ва­тель­ные дво­ич­ные числа (от 00 до 11, со­от­вет­ствен­но). За­ко­ди­руй­те таким об­ра­зом по­сле­до­ва­тель­ность сим­во­лов ББГА и за­пи­ши­те по­лу­чен­ное дво­ич­ное чис­ло в шест­на­дца­те­рич­ной си­сте­ме счис­ле­ния.

**Решение:**

За­ко­ди­ру­ем по­сле­до­ва­тель­ность букв: ББГА — 01011100. Те­перь разобьём это пред­став­ле­ние на четвёрки спра­ва на­ле­во и пе­ре­ведём по­лу­чен­ный набор чисел сна­ча­ла в де­ся­тич­ный код, затем в шест­на­дца­те­рич­ный:

0101 1100 — 5 12 — 5С.

**Пример №6 с решением**

Черно-белое раст­ро­вое изоб­ра­же­ние ко­ди­ру­ет­ся по­строч­но, на­чи­ная с ле­во­го верх­не­го угла и за­кан­чи­вая в пра­вом ниж­нем углу. При ко­ди­ро­ва­нии 1 обо­зна­ча­ет чер­ный цвет, а 0 – белый. Закодируйте, таким образом, изоб­ра­же­ние и за­пи­ши­те ре­зуль­тат в вось­ме­рич­ной си­сте­ме счис­ле­ния.



**Решение:**

Код пер­вой стро­ки: 10101.

Код вто­рой стро­ки: 11000.

Код тре­тьей стро­ки: 01010.

За­пи­шем коды по по­ряд­ку в одну стро­ку: 101011100001010. Те­перь разобьём это пред­став­ле­ние на трой­ки спра­ва на­ле­во и пе­ре­ведём по­лу­чен­ный набор чисел в де­ся­тич­ный код (вось­ме­рич­ное представление сов­па­да­ет с де­ся­тич­ным при раз­би­е­нии трой­ка­ми).

101 011 100 001 010 — 53412.

**Задания для тренировки**

1) Для ко­ди­ро­ва­ния не­ко­то­рой по­сле­до­ва­тель­но­сти, со­сто­я­щей из букв А, Б, В, Г и Д, ре­ши­ли ис­поль­зо­вать не­рав­но­мер­ный дво­ич­ный код, поз­во­ля­ю­щий од­но­знач­но де­ко­ди­ро­вать дво­ич­ную по­сле­до­ва­тель­ность, по­яв­ля­ю­щу­ю­ся на приёмной сто­ро­не ка­на­ла связи. Для букв А, Б, В и Г ис­поль­зо­ва­ли такие ко­до­вые слова: А–111, Б–110, В–100, Г–101.

Ука­жи­те, каким ко­до­вым сло­вом может быть за­ко­ди­ро­ва­на буква Д. Код дол­жен удо­вле­тво­рять свой­ству од­но­знач­но­го де­ко­ди­ро­ва­ния. Если можно ис­поль­зо­вать более од­но­го ко­до­во­го слова, ука­жи­те крат­чай­шее из них.

1) 0

2) 01

3) 00

4) 000

2) Для кодирования некоторой последовательности, состоящей из букв А, Б, В, Г и Д, используется неравномерный двоичный код, позволяющий однозначно декодировать полученную двоичную последовательность. Вот этот код: А – 00, Б – 01, В – 100, Г – 101, Д – 110. Можно ли сократить для одной из букв длину кодового слова так, чтобы код по-прежнему можно было декодировать однозначно? Коды остальных букв меняться не должны. Выберите правильный вариант ответа.

1) для буквы Д – 11

2) это невозможно

3) для буквы Г – 10

4) для буквы Д – 10

3) По каналу связи передаются сообщения, содержащие только 4 буквы К, О, Р, А; для передачи используется двоичный код, допускающий однозначное декодирование. Для букв Р, А, К используются такие кодовые слова: Р: 000, А: 10, К: 01.

Укажите такое кодовое слово для буквы О, при котором код будет допускать однозначное декодирование. Если таких кодовых слов несколько, укажите то, у которого меньшая длина.

1) 1

2) 0

3) 11

4) 001

4) Для кодирования некоторой последовательности, состоящей из букв У, Ч, Е, Н, И и К, используется неравномерный двоичный префиксный код. Вот этот код: У — 000, Ч — 001, Е — 010, Н — 100, И — 011, К — 11. Можно ли сократить для одной из букв длину кодового слова так, чтобы код по-прежнему остался префиксным? Коды остальных букв меняться не должны.

Выберите правильный вариант ответа.

**Примечание.** Префиксный код — это код, в котором ни одно кодовое слово не является началом другого; такие коды позволяют однозначно декодировать полученную двоичную последовательность.

1) кодовое слово для буквы Е можно сократить до 01;

2) кодовое слово для буквы К можно сократить до 1;

3) кодовое слово для буквы Н можно сократить до 10;

4) это невозможно.

5) Для ко­ди­ро­ва­ния не­ко­то­рой по­сле­до­ва­тель­но­сти, со­сто­я­щей из букв А, Б, В, Г и Д, ре­ши­ли ис­поль­зо­вать не­рав­но­мер­ный дво­ич­ный код, поз­во­ля­ю­щий од­но­знач­но де­ко­ди­ро­вать дво­ич­ную по­сле­до­ва­тель­ность, по­яв­ля­ю­щу­ю­ся на приёмной сто­ро­не ка­на­ла связи. Для букв А, Б, В и Г ис­поль­зо­ва­ли такие ко­до­вые слова: А — 111, Б — 110, В — 101, Г — 100. Ука­жи­те, каким ко­до­вым сло­вом из пе­ре­чис­лен­ных ниже может быть закодирова­на буква Д. Код дол­жен удо­вле­тво­рять свой­ству од­но­знач­но­го де­ко­ди­ро­ва­ния. Если можно ис­поль­зо­вать более од­но­го ко­до­во­го слова, ука­жи­те крат­чай­шее из них.

1) 1

2) 0

3) 01

4) 10

**По­яс­не­ние.**

Для того, чтобы со­об­ще­ние, за­пи­сан­ное с по­мо­щью не­рав­но­мер­но­го по длине кода, од­но­знач­но рас­ко­ди­ро­ва­лось, тре­бу­ет­ся, чтобы ни­ка­кой код не был на­ча­лом дру­го­го (более длин­но­го) кода. Рас­смот­рим ва­ри­ан­ты для буквы Д, на­чи­ная с са­мо­го ко­рот­ко­го.

1) Д=1: код буквы Д яв­ля­ет­ся на­ча­лом всех пред­став­лен­ных кодов букв, по­это­му этот ва­ри­ант не под­хо­дит.

2) Д=0: код буквы Д не яв­ля­ет­ся на­ча­лом дру­го­го кода, по­это­му этот ва­ри­ант под­хо­дит.

3) Д=01: код буквы Д не яв­ля­ет­ся на­ча­лом дру­го­го кода, по­это­му этот ва­ри­ант под­хо­дит.

4) Д=10: код буквы Д яв­ля­ет­ся на­ча­лом кодов букв В и Г, сле­до­ва­тель­но, этот ва­ри­ант не под­хо­дит.

Таким об­ра­зом, под­хо­дят два ва­ри­ан­та: 0 и 01. 0 ко­ро­че, чем 01.

Пра­виль­ный ответ ука­зан под но­ме­ром 2.

Ответ: 2

6) По ка­на­лу связи пе­ре­да­ют­ся со­об­ще­ния, со­дер­жа­щие толь­ко 4 буквы: Е, Н, О, Т.

В любом со­об­ще­нии боль­ше всего букв О, сле­ду­ю­щая по ча­сто­те буква − Е, затем − Н. Буква Т встре­ча­ет­ся реже, чем любая дру­гая. Для пе­ре­да­чи со­об­ще­ний нужно ис­поль­зо­вать не­рав­но­мер­ный дво­ич­ный код, до­пус­ка­ю­щий од­но­знач­ное де­ко­ди­ро­ва­ние; при этом со­об­ще­ния долж­ны быть как можно ко­ро­че. Шиф­ро­валь­щик может ис­поль­зо­вать один из пе­ре­чис­лен­ных ниже кодов. Какой код ему сле­ду­ет вы­брать?

1) Е−0, Н−1, O−00, Т−11

2) O−1, Н−0, Е−01,Т−10

3) Е−1, Н−01, O−001, Т−000

4) О−0, Н−11, Е−101, Т−100

**По­яс­не­ние.**

Вы­бе­рем коды, для ко­то­рых вы­пол­не­но усло­вие Фано. Это коды 3 и 4.

Чтобы со­об­ще­ние было как можно ко­ро­че, не­об­хо­ди­мо, чтобы чем чаще встре­ча­лась буква, тем ко­ро­че был ее код.

Сле­до­ва­тель­но, ответ 4, по­сколь­ку буква О — самая часто встре­ча­ю­ща­я­ся буква и для ее ко­ди­ро­ва­ния в ва­ри­ан­те 4 ис­поль­зу­ет­ся один сим­вол.

Ответ: 4

7) По ка­на­лу связи пе­ре­да­ют­ся со­об­ще­ния, со­дер­жа­щие толь­ко 4 буквы П, О, С, Т; для пе­ре­да­чи ис­поль­зу­ет­ся дво­ич­ный код, до­пус­ка­ю­щий од­но­знач­ное де­ко­ди­ро­ва­ние. Для букв Т, О, П ис­поль­зу­ют­ся такие ко­до­вые слова: Т: 111, О: 10, П: 01. Ука­жи­те такое ко­до­вое слово для буквы С, при ко­то­ром код будет до­пус­кать од­но­знач­ное де­ко­ди­ро­ва­ние. Если таких ко­до­вых слов не­сколь­ко, ука­жи­те тот, у ко­то­ро­го мень­шая длина.

1) 1

2) 0

3) 00

4) 110

8) Для ко­ди­ро­ва­ния со­об­ще­ния, со­сто­я­ще­го толь­ко из букв A, B, C, D и E, ис­поль­зу­ет­ся не­рав­но­мер­ный по длине дво­ич­ный код:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A | B | C | D | E |
| 000 | 11 | 01 | 001 | 10 |

Какое (толь­ко одно!) из че­ты­рех по­лу­чен­ных со­об­ще­ний было пе­ре­да­но без оши­бок и может быть рас­ко­ди­ро­ва­но:

1) 110000010011110

2) 110000011011110

3) 110001001001110

4) 110000001011110

9) Для ко­ди­ро­ва­ния букв О, К, Г, Д, Р ре­ши­ли ис­поль­зо­вать дво­ич­ное пред­став­ле­ние чисел 0, 1, 2, 3 и 4 со­от­вет­ствен­но (с со­хра­не­ни­ем од­но­го не­зна­ча­ще­го нуля в слу­чае од­но­раз­ряд­но­го пред­став­ле­ния). За­ко­ди­руй­те по­сле­до­ва­тель­ность букв ГО­РО­ДОК таким спо­со­бом и ре­зуль­тат за­пи­ши­те вось­ме­рич­ным кодом.

10) Для пе­ре­да­чи по ка­на­лу связи со­об­ще­ния, со­сто­я­ще­го толь­ко из сим­во­лов А, Б, В и Г, ис­поль­зу­ет­ся не­рав­но­мер­ный (по длине) код: А-10, Б-11, В-001, Г-011. Через канал связи пе­ре­да­ет­ся со­об­ще­ние: АБ­ГВГБ. За­ко­ди­руй­те со­об­ще­ние дан­ным кодом. По­лу­чен­ное дво­ич­ное число пе­ре­ве­ди­те в шест­на­дца­те­рич­ный вид.

**Ответы к заданиям для тренировки**

1) 1

2) 1

3) 3

4) 3

5) 2

6) 4

7) 3

8) 1

9) 42061

10) 5B2F.