# 5 (базовый уровень, время – 2 мин)

**Тема**: Кодирование и декодирование информации.

**Что нужно знать**:

* кодирование – это перевод информации с одного языка на другой (запись в другой системе символов, в другом алфавите)
* обычно кодированием называют перевод информации с «человеческого» языка на формальный, например, в двоичный код, а декодированием – обратный переход
* один символ исходного сообщения может заменяться одним символом нового кода или несколькими символами, а может быть и наоборот – несколько символов исходного сообщения заменяются одним символом в новом коде (китайские иероглифы обозначают целые слова и понятия)
* кодирование может быть *равномерное* и *неравномерное*;при равномерном кодировании все символы кодируются кодами равной длины;
при неравномерном кодировании разные символы могут кодироваться кодами разной длины, это затрудняет декодирование
* закодированное сообщение можно однозначно декодировать с начала, если выполняется *условие Фано*: никакое кодовое слово не является началом другого кодового слова;
* закодированное сообщение можно однозначно декодировать с конца, если выполняется *обратное условие Фано*: никакое кодовое слово не является окончанием другого кодового слова;
* условие Фано – это достаточное, но не необходимое условие однозначного декодирования.

### Пример задания

**Р-14.** Для кодирования некоторой последовательности, состоящей из букв А, Б, В, Г, Д, Е, решили использовать неравномерный двоичный код, удовлетворяющий условию Фано. Для буквы А использовали кодовое слово 0; для буквы Б – кодовое слово 10. Какова наименьшая возможная

сумма длин всех шести кодовых слов? Примечание. Условие Фано означает, что никакое кодовое слово не является началом другого кодового слова. Это обеспечивает возможность однозначной расшифровки закодированных сообщений.

**Решение**:

1. это задание удобнее решать с помощью дерева; условие Фано выполняется тогда, когда все выбранные кодовые слова заканчиваются в листьях дерева
2. построим дерево по известным кодовым словам: А – 0, Б – 10:

1

0

1

0

1. на оставшуюся свободную ветку нужно «повесить» 4 кодовых слова (для букв В, Г, Д, Е)
2. если выбрать один код длиной 3 (В – 110), то оставшиеся 3 кода нужно «повесить» на одну ветку, так, что на ней нужно делать две развилки:

1

0

1

0

0

1

1

0

0

1

1. суммарная длина кодовых слов будет в этом случае равна

1 + 2 + 3 + 4 + 2·5 = 20

1. попробуем другой вариант: оставшиеся 4 кода повесить на 4 ветки одинаковой длины:

1

0

1

0

1

0

1

0

1

0

1. суммарная длина кодовых слов будет в этом случае меньше, чем в предыдущем случае:

1 + 2 + 4·4 = 19

1. Ответ: 19.

### Ещё пример задания

**Р-13.** *По каналу связи передаются сообщения, каждое из которых содержит 16 букв А, 8 букв Б, 4 буквы В и 4 буквы Г (других букв в сообщениях нет). Каждую букву кодируют двоичной последовательностью. При выборе кода учитывались два требования:*

 *а) ни одно кодовое слово не является началом другого (это нужно, чтобы код допускал однозначное декодирование);*

 *б) общая длина закодированного сообщения должна быть как можно меньше.*

*Какой код из приведённых ниже следует выбрать для кодирования букв А, Б, В и Г?*

*1) А:0, Б:10, В:110, Г:111*

*2) А:0, Б:10, В:01, Г:11*

*3) А:1, Б:01, В:011, Г:001*

*4) А:00, Б:01, В:10, Г:11*

**Решение**:

1. сначала выберем коды, в которых ни одно кодовое слово не совпадет с началом другого (такие коды называю префиксными)
2. для кода 2 условие «а» не выполняется, так как кодовое слово буквы В (01) начинается с кодового слова буквы А (0)
3. для кода 3 условие «а» не выполняется, так как кодовое слово буквы В (011) начинается с кодового слова буквы Б (01)
4. для кодов 1 и 4 условие выполняется, их рассматриваем дальше
5. считаем общее количество битов в сообщении для кода 1:

16∙1 + 8·2 + 4∙3 + 4∙3 = 56 битов

1. считаем общее количество битов в сообщении для кода 4:

16∙2 + 8·2 + 4∙2 + 4∙2 = 64 бита

1. код 1 даёт наименьшую длину сообщения, поэтому выбираем его
2. Ответ: 1.

### Ещё пример задания

**Р-12.** *Для кодирования некоторой последовательности, состоящей из букв А, Б, В, Г, решили использовать неравномерный двоичный код, удовлетворяющий условию Фано. Для буквы А использовали кодовое слово 0, для буквы Б – кодовое слово 110. Какова наименьшая возможная суммарная длина всех четырёх кодовых слов?*

1) 7 2) 8 3) 9 4) 10

**Решение (способ 1, исключение вариантов)**:

1. условие Фано означает, что ни одно кодовое слово не совпадает с началом другого кодового слова
2. поскольку уже есть кодовое слово 0, ни одно другое кодовое слово не может начинаться с 0
3. поскольку есть код 110, запрещены кодовые слова 1, 11; кроме того, ни одно другое кодовое слово не может начинаться с 110
4. таким образом, нужно выбрать еще два кодовых слова, для которых выполняются эти ограничения
5. есть одно допустимое кодовое слово из двух символов: 10
6. если выбрать кодовое слово 10 для буквы В, то остаётся одно допустимое трёхсимвольное кодовое слово – 111, которое можно выбрать для буквы Г
7. таким образом, выбрав кодовые слова А – 0, Б – 110, В – 10, Г – 111, получаем суммарную длину кодовых слов 9 символов
8. если же не выбрать В – 10, то есть три допустимых трёхсимвольных кодовых слова: 100, 101 и 110; при выборе любых двух их них для букв В и Г получаем суммарную длину кодовых слов 10, что больше 9; поэтому выбираем вариант 3 (9 символов)
9. Ответ: 3.

**Решение (способ 2, построение дерева)**:

1. условие Фано означает, что ни одно кодовое слово не совпадает с началом другого кодового слова; при этом в дереве кода все кодовые слова должны располагаться в листьях дерева, то есть в узлах, которые не имеют потомков;
2. построим дерево для заданных кодовых слов А – 0 и Б – 110:

1

0

1

0

0

1. штриховыми линиями отмечены две «пустые» ветви, на которые можно «прикрепить» листья для кодовых слов букв В (10) и Г (111)

1

0

1

0

0

1. таким образом, выбрав кодовые слова А – 0, Б – 110, В – 10, Г – 111, получаем суммарную длину кодовых слов 9 символов
2. Ответ: 3.

### Ещё пример задания

**Р-11.** *По каналу связи передаются сообщения, содержащие только 5 букв А, И, К, О, Т. Для кодирования букв используется неравномерный двоичный код с такими кодовыми словами:*

*А — 0, И — 00, К — 10, О — 110, Т — 111.*

*Среди приведённых ниже слов укажите такое, код которого можно декодировать только одним способом. Если таких слов несколько, укажите первое по алфавиту.*

1) КАА 2) ИКОТА 3) КОТ 4) ни одно из со­об­ще­ний не под­хо­дит

**Решение**:

1. прежде всего заметим, что для заданного кода не выполняется ни прямое, ни обратное условие Фано; «виновата» в этом пара А – И: код буквы А совпадает как с началом, так и с окончанием кода буквы И; больше ни для одной пары кодовых слов прямое условие Фано не нарушено
2. это означает, что не все сообщения могут быть декодированы однозначно
3. теперь нужно понять, какие последовательности могут быть декодированы неоднозначно; в данном случае очевидно, что сообщения АА и И кодируются одинаково: 00, поэтому все слова, где есть АА или И, не могут быть декодированы однозначно
4. поэтому варианты 1 (КАА) и 2 (ИКОТА) отпадают
5. на всякий случай проверим вариант 3: КОТ = 10110111; первой буквой может быть только К (по-другому сочетание 10 получить нельзя), аналогично вторая буква – только О, а третья – только Т
6. Ответ: 3.

### Ещё пример задания

**Р-10.** *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 (способ 1, исключение вариантов)**:

1. код однозначно декодируется, если выполняется условие Фано или обратное условие Фано; в данном случае «прямое» условие Фано выполняется: с кода буквы О (0) не начинается ни один из двух других кодов;
2. новый код не может начинаться с нуля (иначе нарушится условие Фано)
3. начнём проверку с кодов длиной 1; единственный код, не начинающийся с нуля – 1 – не подходит, потому что с 1 начинаются два других кода: Т (111) и П (100
4. кодов длиной 2, начинающихся с 1, всего 2: 10 и 11, но их использовать нельзя, потому что с 10 начинается код буквы П, а с 11 – код буквы Т
5. рассматриваем коды длиной 3, начинающиеся с 1; коды 100 и 111 уже заняты, а ещё два – 101 и 110 – свободны и их можно использовать, причём условие Фано выполняется в обоих случаях;
6. поскольку нужно выбрать код с минимальным значением, выбираем 101
7. Ответ: 101.

**Решение (способ 2, построение дерева)**:

1. условие Фано означает, что ни одно кодовое слово не совпадает с началом другого кодового слова; при этом в дереве кода все кодовые слова должны располагаться в листьях дерева, то есть в узлах, которые не имеют потомков;
2. 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. таким образом, выбрав кодовые слова А – 0, Б – 110, В – 10, Г – 111, получаем суммарную длину кодовых слов 9 символов
2. Ответ: 101.

### Ещё пример задания

**Р-09.** *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 – 0; Б – 100; В – 1010; Г – 111; Д – 110. Требуется сократить для одной из букв длину кодового слова так, чтобы код по-прежнему можно было декодировать однозначно. Коды остальных букв меняться не должны.*

*Каким из указанных способов это можно сделать?*

1) для буквы В – 101 2) это невозможно

3) для буквы В – 010 4) для буквы Б – 10

**Решение**:

1. код однозначно декодируется, если выполняется условие Фано или обратное условие Фано; в данном случае «прямое» условие Фано выполняется: с кода буквы А (0) не начинается ни один другой код, оставшиеся короткие коды (Б, Г и Д) не совпадают с началом длинного кода буквы В; таким образом, при сокращении нужно сохранить выполнение условия Фано
2. вариант 3 не подходит, потому что новый код буквы В начинается с 0 (кода А), поэтому условие Фано нарушено
3. вариант 4 не подходит, потому что код буквы В начинается с 10 (нового кода б), поэтому условие Фано нарушено
4. вариант 1 подходит, условие Фано сохраняется (все трёхбитные коды различны, ни один не начинается с 0)
5. Ответ: 1.

### Ещё пример задания

**Р-08.** *По каналу связи передаются сообщения, содержащие только 4 буквы: А, И, С, Т.*

*В любом сообщении больше всего букв А, следующая по частоте буква – С, затем – И. 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) А – 0, И – 1, С – 00, Т – 11 2) С – 1, И – 0, А – 01, Т – 10

3) А – 1, И – 01, С – 001, Т – 000 4) С – 0, И – 11, А – 101, Т – 100

**Решение**:

1. сначала выберем коды, допускающие однозначное декодирование: это коды 3 и 4 (для них выполняется условие Фано), коды 1 и 2 не подходят
2. для того, чтобы длина сообщения была как можно короче, должно выполнять правило: «чем чаще встречается буква, тем короче её код»;
3. к сожалению, правило, приведённое выше, не совсем «хорошо» выполняется для кодов 3 и 4: в коде 3 длина кодового слова для буквы С больше, чем длина кодового слова буквы И (а хочется наоборот); для кода 4 длина кодового слова для буквы А – не самая маленькая из всех
4. сравним коды 3 и 4, предполагая, что в сообщении буква А встречается α раз, буква С – β раз, буква И – γ раз и буква Т – δ раз; причём по условию задачи α > β > γ > δ
5. при кодировании кодом 3 получаем сообщение длиной

L3  = α + 3β + 2γ +3 δ

1. при кодировании кодом 4 получаем сообщение длиной

L4  = 3α + β + 2γ +3 δ

1. находим разность: L4 ­– L3 = (3α + β + 2γ +3 δ) – (α + 3β + 2γ +3 δ) = 2α – 2β
2. поскольку α > β, получаем L4 ­– L3 > 0, то есть код 3 более экономичный
3. Ответ: 3.

### Ещё пример задания

**Р-07.** *По каналу связи передаются сообщения, содержащие только 4 буквы: Е, Н, О, Т. Для кодирования букв Е, Н, О используются 5-битовые кодовые слова: Е - 00000, Н - 00111, О - 11011. Для этого набора кодовых слов выполнено такое свойство: любые два слова из набора отличаются не менее чем в трёх позициях. Это свойство важно для расшифровки сообщений при наличии помех. Какое из перечисленных ниже кодовых слов можно использовать для буквы Т, чтобы указанное свойство выполнялось для всех четырёх кодовых слов?*

1) 11111 2) 11100 3) 00011 4) не подходит ни одно из указанных выше слов

**Решение**:

1. код, рассмотренный в условии задачи, относится к помехоустойчивым кодам, которые позволяют обнаружить и исправить определенное количество ошибок, вызванных помехами при передаче данных;
2. количество позиций, в которых отличаются два кодовых слова одинаковой длины, называется расстоянием Хэмминга
3. код, в котором расстояние Хэмминга между каждой парой кодовых слов равно *d*, позволяет обнаружить до *d-1* ошибок; для исправления *r* ошибок требуется выполнение условия

*d ≥ 2r + 1*

поэтому код с *d = 3* позволяет обнаружить одну или две ошибки, и исправить одну ошибку.

1. легко проверить, что для заданного кода (Е - 00000, Н - 00111, О - 11011) расстояние Хэмминга равно 3; в таблице выделены отличающиеся биты, их по три в парах Е-Н и Н-О и четыре в паре Е-О:

**Е – 00000 Е – 00000 Н – 00111**

**Н – 00111 О – 11011 О – 11011**

1. теперь проверяем расстояние между известными кодами и вариантами ответа; для первого ответа 11111 получаем минимальное расстояние 1 (в паре О-Т), этот вариант не подходит:

**Е – 00000 Н – 00111 О – 11011**

**Т - 11111 Т - 11111 Т - 11111**

1. для второго ответа 11100 получаем минимальное расстояние 3 (в парах Е-Т и О-Т):

**Е – 00000 Н – 00111 О – 11011**

**Т - 11100 Т - 11100 Т - 11100**

1. для третьего ответа 00011 получаем минимальное расстояние 1 (в паре Н-Т) , этот вариант не подходит:

**Е – 00000 Н – 00111 О – 11011**

**Т - 00011 Т - 00011 Т - 00011**

1. таким образом, расстояние Хэмминга, равное 3, сохраняется только для ответа 2
2. Ответ: 2.

### Ещё пример задания:

**Р-06.** *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–010, В–011, Г–101, Д–111. 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 Б – 01 2) это невозможно

3) для буквы В – 01 4) для буквы Г – 01

**Решение (1 способ, проверка условий Фано)**:

1. для однозначного декодирования достаточно, чтобы выполнялось условие Фано или обратное условие Фано;
2. проверяем последовательно варианты 1, 3 и 4; если ни один из них не подойдет, придется выбрать вариант 2 («это невозможно»);
3. проверяем вариант 1: *А–00, Б–01, В–011, Г–101, Д–111.*

«прямое» условие Фано не выполняется (код буквы Б совпадает с началом кода буквы В);

«обратное» условие Фано не выполняется (код буквы Б совпадает с окончанием кода буквы Г); поэтому этот вариант не подходит;

1. проверяем вариант 3: *А–00, Б–010, В–01, Г–101, Д–111.*

«прямое» условие Фано не выполняется (код буквы В совпадает с началом кода буквы Б);

«обратное» условие Фано не выполняется (код буквы В совпадает с окончанием кода буквы Г); поэтому этот вариант не подходит;

1. проверяем вариант 4: *А–00, Б–010, В–011, Г–01, Д–111.*

«прямое» условие Фано не выполняется (код буквы Г совпадает с началом кодов букв Б и В); но **«обратное» условие Фано выполняется** (код буквы Г не совпадает с окончанием кодов остальных буквы); поэтому этот вариант подходит;

1. правильный ответ – 4.

**Решение (2 способ, дерево)**:

1. построим двоичное дерево, в котором от каждого узла отходит две ветки, соответствующие выбору следующей цифры кода – 0 или 1; разместим на этом дереве буквы А, Б, В, Г и Д так, чтобы их код получался как последовательность чисел на рёбрах, составляющих путь от корня до данной буквы (красным цветом выделен код буквы В – 011):

**0**

0

0

0

0

1

0

1

1

**1**

**1**

0

1

1

корень

1. здесь однозначность декодирования получается за счёт того, что при движении от корня к любой букве в середине пути не встречается других букв (выполняется условие Фано);
2. теперь проверим варианты ответа: предлагается перенести одну из букв, Б, В или Г, в узел с кодом 01, выделенный синим цветом
3. видим, что при переносе любой из этих букв нарушится условие Фано; например, при переносе буквы Б в синий узел она оказывается на пути от корня до В, и т.д.; это значит, что предлагаемые варианты не позволяют выполнить прямое условие Фано
4. хочется уже выбрать вариант 2 («это невозможно»), но у нас есть еще обратное условие Фано, для которого тоже можно построить аналогичное дерево, в котором движение от корня к букве дает её код **с конца** (красным цветом выделен код буквы В – 011, записанный с конца):

0

0

0

0

0

1

**0**

1

1

1

1

0

**1**

**1**

корень

видно, что обратное условие Фано также выполняется, потому что на пути от корня к любой букве нет других букв

1. в заданных вариантах ответа предлагается переместить букву Б, В или Г в синий узел; понятно, что Б или В туда перемещать нельзя – перемещённая буква отказывается на пути от корня к букве Г; а вот букву Г переместить можно, при этом обратное условие Фано сохранится
2. правильный ответ – 4.

### Ещё пример задания:

**Р-05.** *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. Использовали код:
А–1, Б–000, В–001, Г–011. Укажите, каким кодовым словом должна быть закодирована буква Д. Длина этого кодового слова должна быть наименьшей из всех возможных. Код должен удовлетворять свойству однозначного декодирования.*

1) 00 2) 01 3)11 4) 010

**Решение**:

1. заметим, что для известной части кода выполняется условие Фано – никакое кодовое слово не является началом другого кодового слова
2. если Д = 00, такая кодовая цепочка совпадает с началом Б = 000 и В = 001, невозможно однозначно раскодировать цепочку 000000: это может быть ДДД или ББ; поэтому первый вариант не подходит
3. если Д = 01, такая кодовая цепочка совпадает с началом Г = 011, невозможно однозначно раскодировать цепочку 011: это может быть ДА или Г; поэтому второй вариант тоже не подходит
4. если Д = 11, условие Фано тоже нарушено: кодовое слово А = 1 совпадает с началом кода буквы Д, невозможно однозначно раскодировать цепочку 111: это может быть ДА или ААА; третий вариант не подходит
5. для четвертого варианта, Д = 010, условие Фано не нарушено;
6. правильный ответ – 4.

|  |
| --- |
| **Возможные ловушки**:* условие Фано – это **достаточное**, но не необходимое условие однозначного декодирования, поэтому для уверенности полезно найти для всех «неправильных» вариантов контрпримеры: цепочки, для которых однозначное декодирование невозможно
 |

### Еще пример задания:

**Р-04.** *Для кодирования букв А, Б, В, Г решили использовать двухразрядные последовательные двоичные числа (от 00 до 11, соответственно). Если таким способом закодировать последовательность символов БАВГ и записать результат шестнадцатеричным кодом, то получится*

1) 4B16 2) 41116 3)BACD16 4) 102316

**Решение**:

1. из условия коды букв такие: A – 00, Б –01, В – 10 и Г – 11, код равномерный
2. последовательность БАВГ кодируется так: 01 00 10 11 = 1001011
3. разобьем такую запись на тетрады справа налево и каждую тетраду переведем в шестнадцатеричную систему (то есть, сначала в десятичную, а потом заменим все числа от 10 до 15 на буквы A, B, C, D, E, F); получаем

1001011 = 0100 10112 = 4B16

1. правильный ответ – 1.

|  |
| --- |
| **Возможные ловушки**:* расчет на то, что при переводе тетрад в шестнадцатеричную систему можно забыть заменить большие числа (10–15) на буквы (10112 = 11, получаем неверный ответ 41116)
* может быть дан неверный ответ, в котором нужные цифры поменяли местами (расчет на невнимательность), например, B416
* в ответах дана последовательность, напоминающая исходную (неверный ответ BACD16), чтобы сбить случайное угадывание
 |

### Еще пример задания:

**Р-03.** *Для 5 букв латинского алфавита заданы их двоичные коды (для некоторых букв – из двух бит, для некоторых – из трех). Эти коды представлены в таблице:*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A | B | C | D | E |
| 000 | 01 | 100 | 10 | 011 |

Определить, какой набор букв закодирован двоичной строкой 0110100011000

1) EBCEA 2) BDDEA 3) BDCEA 4) EBAEA

 **Решение (вариант 1, декодирование с начала)**:

1. здесь используется неравномерное кодирование, при котором декодирование может быть неоднозначным, то есть, заданному коду может соответствовать несколько разных исходных сообщений
2. попробуем декодировать с начала цепочки, первой буквой может быть B или E, эти случаи нужно рассматривать отдельно
3. пусть первая буква – E с кодом 011, тогда остается цепочка 0100011000
* для кода 0100011000 первой буквой может быть только B с кодом 01, тогда остается 00011000 ( начало исходной цепочки – EB?)
* для кода 00011000 первой буквой может быть только A с кодом 000, тогда остается 11000, а эта цепочка не может быть разложена на заданные коды букв
* поэтому наше предположение о том, что первая буква – E, неверно
1. пусть первая буква – B с кодом 01, тогда остается цепочка 10100011000
* для кода 10100011000 первой буквой может быть только D с кодом 10, тогда остается 100011000 (можно полагать, что начало исходной цепочки – BD?)
* для кода 100011000 первой буквой может быть только С с кодом 100, тогда остается 011000 (начало исходной цепочки – BDC?)

Несмотря на то, что среди ответов есть единственная цепочка, которая начинается с BDC, здесь нельзя останавливаться, потому что «хвост» цепочки может «не сойтись»

* для кода 011000 на первом месте может быть B (код 01) или E (011); в первом случае «хвост» 1000 нельзя разбить на заданные коды букв, а во втором – остается код 000 (буква А), поэтому исходная цепочка может быть декодирована как BDCEA
1. правильный ответ – 3

|  |
| --- |
| **Возможные ловушки и проблемы**:* при декодировании неравномерных кодов может быть очень много вариантов, их нужно рассмотреть все; это требует серьезных усилий и можно легко запутаться
* нельзя останавливаться, не закончив декодирование до конца и не убедившись, что все «сходится», на это обычно и рассчитаны неверные ответы
 |

**Решение (вариант 2, декодирование с конца)**:

1. для кода 0110100011000 последней буквой может быть только А (код 000), тогда остается цепочка 0110100011
2. для 0110100011 последней может быть только буква E (011), тогда остается цепочка 0110100
3. для 0110100 последней может быть только буква C (100), тогда остается цепочка 0110
4. для 0110 последней может быть только буква D (10), тогда остается 01 – это код буквы B
5. таким образом, получилась цепочка BDCEA
6. правильный ответ – 3

|  |
| --- |
| **Возможные ловушки и проблемы**:* при декодировании неравномерных кодов может быть очень много вариантов (здесь *случайно* получилась единственно возможная цепочка), их нужно рассмотреть все; это требует серьезных усилий и можно легко запутаться
* нельзя останавливаться, не закончив декодирование до конца и не убедившись, что все «сходится», на это обычно и рассчитаны неверные ответы
 |

**Решение (вариант 3, кодирование ответов)**:

1. в данном случае самое простое и надежное – просто закодировать все ответы, используя приведенную таблицу кодов, а затем сравнить результаты с заданной цепочкой
2. получим

1) EBCEA – 01101100011000 2) BDDEA – 011010011000

3) BDCEA – 0110100011000 4) EBAEA – 01101000011000

1. сравнивая эти цепочки с заданной, находим, что правильный ответ – 3.

|  |
| --- |
| **Возможные проблемы**:* сложно сравнивать длинные двоичные последовательности, поскольку они однородны, содержат много одинаковых нулей и единиц
 |

### Еще пример задания:

**Р-02.** *Для передачи по каналу связи сообщения, состоящего только из букв А, Б, В, Г, решили использовать неравномерный по длине код: A=0, Б=10, В=110. Как нужно закодировать букву Г, чтобы длина кода была минимальной и допускалось однозначное разбиение кодированного сообщения на буквы?*

1) 1 2) 1110 3) 111 4) 11

**Решение (вариант 1, метод подбора)**:

1. рассмотрим все варианты в порядке увеличения длины кода буквы Г
2. начнем с Г=1; при этом получается, что сообщение «10» может быть раскодировано двояко: как ГА или Б, поэтому этот вариант не подходит
3. следующий по длине вариант – Г=11; в этом случае сообщение «110» может быть раскодировано как ГА или В, поэтому этот вариант тоже не подходит
4. третий вариант, Г=111, дает однозначное раскодирование во всех сочетаниях букв, поэтому…
5. … правильный ответ – 3.

|  |
| --- |
| **Возможные проблемы**:* при переборе можно ошибиться и «просмотреть» какой-нибудь вариант
 |

**Решение (вариант 2, «умный» метод)**:

1. для того, чтобы сообщение, записанное с помощью неравномерного по длине кода, однозначно раскодировалось, требуется, чтобы никакой код не был началом другого (более длинного) кода; это условие называют *условием Фано*
2. как и в первом решении, рассматриваем варианты, начиная с самого короткого кода для буквы Г; в нашем случае код Г=1 является началом кодов букв Б и В, поэтому условие Фано не выполняется, такой код не подходит
3. код Г=11 также является началом другого кода (кода буквы В), поэтому это тоже ошибочный вариант
4. третий вариант кода, Г=111, не является началом никакого уже известного кода; кроме того, ни один уже имеющийся код не является началом кода 111; таким образом, условие Фано выполняется
5. поэтому правильный ответ – 3.

|  |
| --- |
| **Возможные проблемы**:* нужно знать условие Фано
 |

### Еще пример задания[[1]](#footnote-1):

**Р-01.** *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.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

*Для компактности результат записали в шестнадцатеричной системе счисления. Выберите правильную запись кода.*

1) BD9AA5 2) BDA9B5 3) BDA9D5 4) DB9DAB

**Решение**:

1. «вытянем» растровое изображение в цепочку: сначала первая (верхняя) строка, потом – вторая, и т.д.:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1 строка* | *2 строка* | *3 строка* | *4 строка* |

1. в этой полоске 24 ячейки, черные заполним единицами, а белые – нулями:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *1* | *0* | *1* | *1* | *1* | *1* | *0* | ***1*** | ***1*** | *0* | ***1*** | *0* | ***1*** | *0* | *0* | ***1*** | ***1*** | ***1*** | *0* | ***1*** | *0* | ***1*** | *0* | ***1*** |
| *1 строка* | *2 строка* | *3 строка* | *4 строка* |

1. поскольку каждая цифра в шестнадцатеричной системе раскладывается ровно в 4 двоичных цифры, разобьем полоску на **тетрады** – группы из четырех ячеек (в данном случае все равно, откуда начинать разбивку, поскольку в полоске целое число тетрад – 6):

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *1* | *0* | *1* | *1* | *1* | *1* | *0* | ***1*** | ***1*** | *0* | ***1*** | *0* | ***1*** | *0* | *0* | ***1*** | ***1*** | ***1*** | *0* | ***1*** | *0* | ***1*** | *0* | ***1*** |

1. переводя тетрады в шестнадцатеричную систему, получаем последовательно цифры B (11), D(13), A(10), 9, D(13) и 5, то есть, цепочку BDA9D5
2. поэтому правильный ответ – 3.

|  |
| --- |
| **Возможные проблемы**:* нужно уметь быстро переводить тетрады в шестнадцатеричные цифры (в крайнем случае, это можно сделать через десятичную систему)
 |

### Еще пример задания:

**Р-00.** *Для передачи чисел по каналу с помехами используется код проверки четности. Каждая его цифра записывается в двоичном представлении, с добавлением ведущих нулей до длины 4, и к получившейся последовательности дописывается сумма её элементов по модулю* **2** *(например, если передаём* **23***, то получим последовательность* **0010100110***). Определите, какое число передавалось по каналу в виде* **01010100100111100011***?*

1) 59143 2) 5971 3) 102153 4) 10273

**Решение**:

1. сначала разберемся, как закодированы числа в примере; очевидно, что используется код равномерной длины; поскольку 2 знака кодируются 10 двоичными разрядами (битами), на каждую цифру отводится 5 бит, то есть

2 → 00101 и 3 → 00110

1. как следует из условия, четыре первых бита в каждой последовательности – это двоичный код цифры, а пятый бит (бит четности) используется для проверки и рассчитывается как «сумма по модулю два», то есть остаток от деления суммы битов на 2; тогда

2 = 00102, бит четности (0 + 0 + 1 + 0) mod 2 = 1

3 = 00112, бит четности (0 + 0 + 1 + 1) mod 2 = 0

1. но бит четности нам совсем **не нужен**, важно другое: пятый бит в каждой пятерке **можно отбросить**!
2. разобъем заданную последовательность на группы по 5 бит в каждой:

01010, 10010, 01111, 00011.

1. отбросим пятый (последний) бит в каждой группе:

0101, 1001, 0111, 0001.
это и есть двоичные коды передаваемых чисел:

01012 = 5, 10012 = 9, 01112 = 7, 00012 = 1.

1. таким образом, были переданы числа 5, 9, 7, 1 или число 5971.
2. Ответ: 2.
1. Самылкина Н.Н., Островская Е.М. Информатика: тренировочные задания. – М.: Эксмо, 2009. [↑](#footnote-ref-1)