

Информатика. Методическое пособие по задаче №4.

Содержание

1	Введение	2
2	Теория	3
2.1	Перевод в двоичную систему счисления	3
2.2	Перевод из двоичной систему счисления в десятичную	3
2.3	Двоичное дерево	4
2.4	Условие Фано	4
2.5	Обратное условие Фано	5
3	Решение основных прототипов	6
3.1	Кодирование буквенных сообщений в двоичный код	6
3.1.1	Задача 1	6
3.1.2	Задача 2	7
3.2	Закодированы все буквы алфавита	8
3.2.1	Задача 3	8
3.3	Обратное условие Фано	10
3.3.1	Задача 4	10
4	Заключение	11

1 Введение

4-й номер ЕГЭ по информатике посвящен кодированию и декодированию. Для решения этого номера Вам необходимо уметь переводить числа в двоичную систему счисления и из нее.

В данной методичке есть лишь напоминание как переводить числа в двоичную систему счисления и обратно, если вы еще не знакомы с различными системами счисления рекомендуем сначала посмотреть этот ролик: **Системы счисления**.

В этой методичке также будут рассмотрены все существующие прототипы 4-го номера.

2 Теория

2.1 Перевод в двоичную систему счисления

Для начала научимся переводить из десятичной СС в двоичную. Для этого рассмотрим пример перевода из десятичной в двоичную числа 23.

Для того, чтобы это сделать, необходимо разделить нацело исходное число на 2, а также запомнить остаток.

Шаг 1: $23 / 2 = 11$ (ост. 1)

После этого, полученное частное нужно снова поделить на 2. Так делим до тех пор, пока частное не станет равно 0.

Шаг 2: $11 / 2 = 5$ (ост. 1)

Шаг 3: $5 / 2 = 2$ (ост. 1)

Шаг 4: $2 / 2 = 1$ (ост. 0)

Шаг 5: $1 / 2 = 0$ (ост. 1)

После этого необходимо собрать наше число. Для этого записываем остатки от деления, начиная с последнего: 10111_2 – число 23 в двоичной системе счисления.

Хитрый способ:

Надо просто понять, из каких степеней 2-ки состоит число, то есть расписать число по степеням 2-ки. Затем собираем число следующим образом: на тех разрядах степени которых присутствуют в разложении ставим 1, на остальных 0 (разряды считаются с 0 справа налево).

Пример: $41 = 32 + 8 + 1 = 2^5 + 2^3 + 2^0$, тогда $41_{10} = 101001_2$.

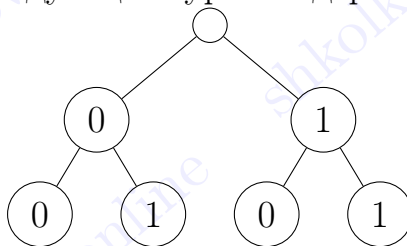
2.2 Перевод из двоичной системы счисления в десятичную

Чтобы перевести из двоичной системы счисления в десятичную, нужно умножить значение каждой цифры на 2 в степени, равной разряду этой цифры (**разряды начинаются с нулевого с правой стороны!**) и полученные значения сложить.

Например: $101010_2 = 2^5 + 2^3 + 2^1 = 32 + 8 + 2 = 42_{10}$

2.3 Двоичное дерево

Двоичное дерево — древовидная структура, все корни которого (**кроме основного**) являются нулями или единицами и имеют по две ветки, которые в свою очередь являются корнями следующего уровня дерева.

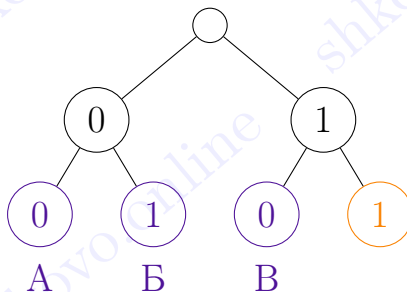


Данное дерево используется при кодировании, его ветви используются в качестве кодов для символов или сообщений.

2.4 Условие Фано

Условие Фано — это свойство бинарного префиксного кода, используемого для кодирования сообщений, которое гласит, что никакое кодовое слово не является префиксом другого кодового слова. Проще говоря, **ни одно кодовое слово не может являться началом другого кодового слова**. Данное условие гарантирует возможность однозначного декодирования двоичного кода (если на вход придет строка кода, то на выход пойдет только один вариант расшифровки).

Условие Фано очень хорошо применяется на двоичном (бинарном) дереве. Допустим, кодовые слова 00, 01, 10 соответствуют буквам А, Б, В. Изобразим это на дереве.



Если потребуется добавить код для какой-нибудь другой буквы, мы не сможем продолжить левые три ветки, так как это нарушит условие Фано и код нельзя будет

декодировать однозначно.

2.5 Обратное условие Фано

Обратное условие Фано — противоположность к прямому условию Фано, **ни одно кодовое слово не может являться концом другого кодового слова**. Данное условие также гарантирует возможность однозначного декодирования двоичного кода.

Обратное условие Фано тоже ложиться на бинарное дерево. Но **коды на дереве должны быть изображены от конца к началу**, то есть код 10 на дереве будет изображен, как 01. Получается, что и коды, которые мы будем дописывать, будут в обратном порядке.

3 Решение основных прототипов

3.1 Кодирование буквенных сообщений в двоичный код

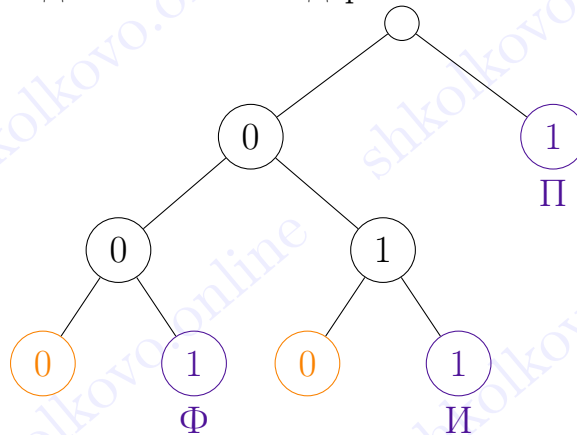
3.1.1 Задача 1

По каналу связи передаются сообщения, содержащие только семь букв: П, И, Ф, А, Г, О, Р; Для передачи используется двоичный код, допускающий однозначное декодирование. Для букв П, И, Ф используются такие кодовые слова: П: 1, И: 011, Ф: 001. Укажите кратчайшее кодовое слово для буквы А, при котором код будет допускать однозначное декодирование. Если таких кодов несколько, укажите код с наименьшим числовым значением.

Примечание. Условие Фано означает, что ни одно кодовое слово не является началом другого кодового слова.

Решение

Изобразим известные кодовые слова на дереве:

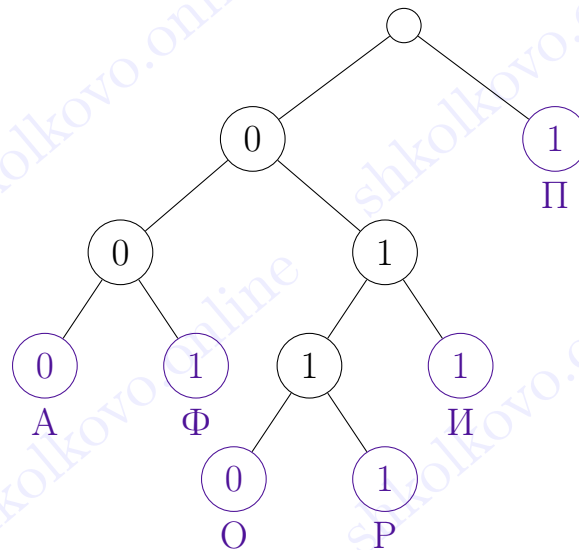


Можно заметить, что на дереве есть две свободные веточки, а букв, которые необходимо закодировать осталось 3. Однако не следует забывать, что бинарное дерево это бесконечная структура, то есть лист 010 можно разделить на два: 0100 и 0111. Аналогично можно сделать и с листом 000. Необходимо выбрать один из этих двух вариантов.

Вспомним, что задаче просят указать **кратчайшее кодовое слово с наименьшим числовым значением для буквы А**. Тогда, верной стратегией будет использовать код 000 для буквы А, а к коду 010 достроить еще ветки.

Тогда ответ на задачу: 000.

Для наглядности построим итоговое дерево (буквы О и Р можно поменять местами, для этой задачи это не играет роли).



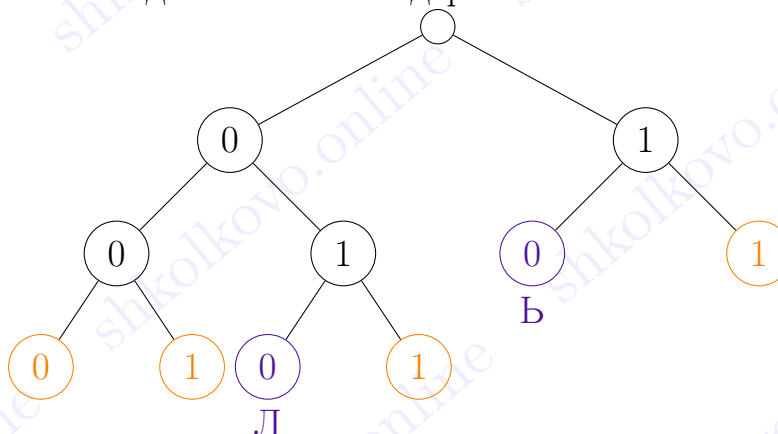
3.1.2 Задача 2

Для кодирования некоторой последовательности, состоящей из букв А, Д, И, Л, Ъ, Н решили использовать неравномерный двоичный код, удовлетворяющий условию Фано. Для букв Л и Ъ использовали кодовые слова 010 и 10 соответственно. Укажите минимальное количество бит, которое потребуется, чтобы закодировать слово ДИАНА?

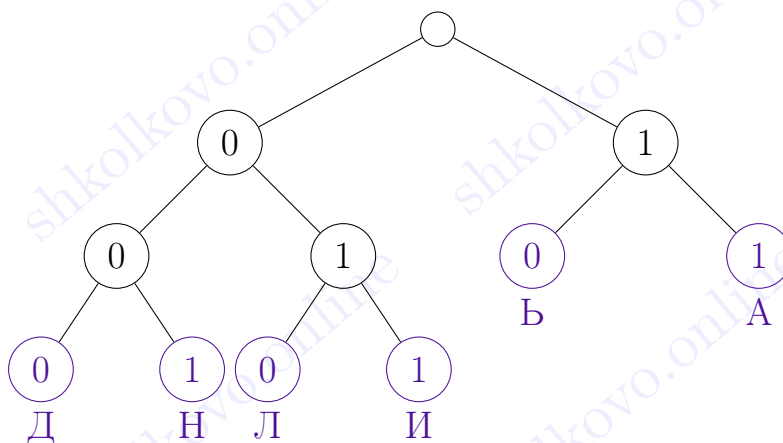
Примечание. Условие Фано означает, что ни одно кодовое слово не является началом другого кодового слова.

Решение

Изобразим известные кодовые слова на дереве:



В условии были даны коды только для 2х из 6 букв. Продлим лист 00, чтобы уместить оставшиеся 4-е буквы. Поскольку в слове ДИАНА два раза повторяется буква А, код для нее следует сделать покороче. Разместим буквы следующим образом.



Совет. Если какая-то буква встречается в слове несколько раз, то код для нее стоит сделать короче, чтобы сэкономить несколько бит.

Теперь нужно лишь посчитать количество бит для кодировки имени ДИАНА, каждый двоичный знак — это 1 бит. **Тогда ответ:** $3 + 3 + 2 + 3 + 2 = 13$.

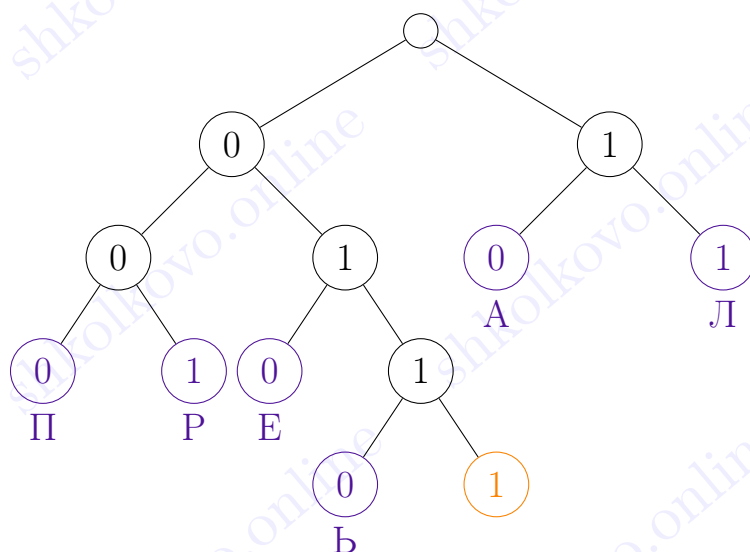
3.2 Закодированы все буквы алфавита

3.2.1 Задача 3

По каналу связи передаются сообщения, содержащие только заглавные буквы русского алфавита. Для передачи используется двоичный код, допускающий однозначное декодирование. Укажите минимальную возможную длину закодированной последовательности ПАРАЛЛЕЛЬ.

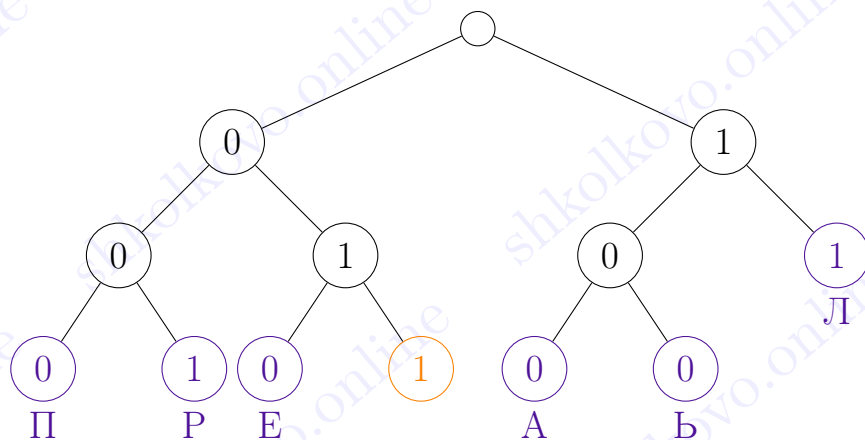
Решение

Нарисуем возможное дерево кодов, используя условие Фано для однозначного декодирования. В слове ПАРАЛЛЕЛЬ буквы А и Л повторяются по несколько раз, попробуем выделить для них слова более короткой длины, чем для остальных букв.



Складываем длины кодов каждой буквы и получаем: $3+2+3+2+2+2+3+2+4 =$
23

Теперь попробуем уменьшить длину Ъ и увеличить длину А



Складываем длины кодов каждой буквы и получаем: $3+3+3+3+2+2+3+2+3 =$
24.

Код стал длиннее, значит нам подходит первый вариант.

Тогда ответ на задачу: 23.

3.3 Обратное условие Фано

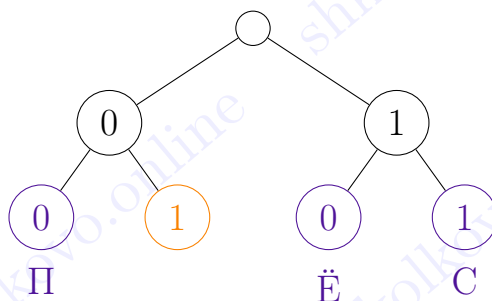
3.3.1 Задача 4

Андрей кодирует буквы П, Ё, С, И, К неравномерным двоичным кодом, который удовлетворяет **обратному** условию Фано. Известно, что букве П соответствует код 00, букве Ё - 01, а букве С - 11. Укажите кратчайшее кодовое слово для буквы И, если таких кодов несколько, то укажите код с наибольшим числовым значением.

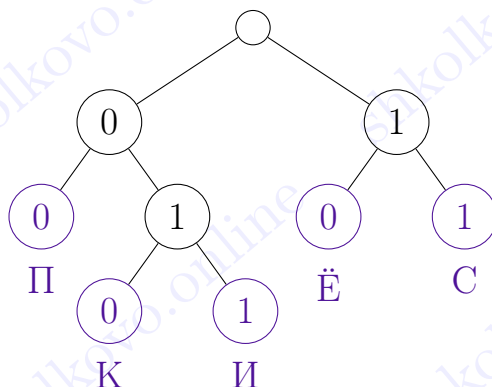
Примечание. Обратное условие Фано означает, что ни одно кодовое слово не является окончанием другого кодового слова.

Решение

Нарисуем «перевернутое» бинарное дерево и разместим буквы, коды которых нам известны.



Заметим, что свободная позиция осталась одна, а буквы, которые необходимо закодировать две. Разветвим лист 01 на 010 и 011. Наибольшее значение имеет лист 011, поэтому используем его для кодирования буквы И.



При использовании обратного условия Фано, полученные с помощью дерева коды букв, необходимо переворачивать. Поэтому **ответ на данную задачу 110**, а не 011.

4 Заключение

Прорезюмируем, что Вы теперь знаете и умеете:

- Умеете переводить числа из десятичной системы счисления и обратно
- Знаете, что такое двоичное дерево и умеете его составлять
- Знаете, что такое условие Фано и обратное условие Фано
- Знаете как кодировать буквенные сообщения
- Умеете решать задачу номер 4 из ЕГЭ

Удачи на экзамене! Ваша команда Школково по Информатике.

*Подробнее рассмотреть тему кодирования вы можете на вебинарах
«Школково»*

А также подписавшись на наши социальные сети:

