

Двоичное кодирование

Когда мы говорим, что человек что-то сообщает, понимает, думает, воспринимает, то используем общее понятие *информации*. Мы говорим, что один человек сообщил информацию другому, а другой её принял, воспринял. Человек что-то увидел, услышал, воспринял зрительную или слуховую информацию.

Компьютер тоже работает с информацией. Эта информация — цепочки нулей и единиц, их называют *двоичные цепочки*. Такие цепочки ещё называются *данными*. Всякую другую, например звуковую, информацию для обработки компьютером надо преобразовать в данные, закодировать в виде двоичной цепочки.

Символы 0 и 1 называются двоичными цифрами или *битами*. Множество этих двух символов называется *двоичным алфавитом*.

Переход от информации к данным — цепочкам в двоичном алфавите — называется *двоичным кодированием*. Двоичным кодированием можно кодировать десятичные числа, тексты на разных языках, изображения и многое другое.

Итак, в нашем распоряжении только нули и единицы. Что можно закодировать этими двумя цифрами? Если использовать в качестве кодов цепочки длины 1, то получим только два кода: 0 и 1. Если использовать цепочки длины 2, то разных кодов будет уже четыре:

00, 01, 10, 11.

Начав кодировать тексты, разработчики компьютеров решили использовать в качестве кодов букв, цифр, знаков препинания и других символов (например, пробела — промежутка между словами) цепочки длины 8 — с их помощью можно закодировать 256 различных символов. Цепочку из восьми нулей и единиц называют *байтом*: один байт состоит из 8 бит.

Почему взяли именно восемь бит? Соображения были разные. 256 байтами можно закодировать все большие и маленькие буквы двух алфавитов — например, латинского и русского, — а также знаки препинания и необходимое количество технических знаков для работы компьютера. Мы обсудим это в задачах.

Сразу скажем, что со временем восьми бит оказалось мало. Это легко было предсказать, ведь человеческих алфавитов много — разных символов много больше, чем 256. Однако поначалу было очень важно всячески экономить память компьютеров, передавать и обрабатывать как можно быстрее тексты, в первую очередь, английские, и поэтому тогда остановились на числе 8.

Термины бит и байт используются при измерении количества информации. Например, цепочка из 24 символов содержит 3 байта информации.

Вот более крупные единицы измерения:

— 1 Кб (килобайт) = 1024 байт

— 1 Мб (мегабайт) = 1024 Кб

— 1 Гб (гигабайт) = 1024 Мб

— 1 Тб (терабайт) = 1024 Гб

Приставка «кило» обычно обозначает умножение на 1000 (километр, килограмм), но в цифровом мире добавление одного бита в цепочку увеличивает количество возможных разных цепочек в 2 раза, поэтому и решили использовать множитель 1024 — самое близкое к 1000 число, получаемое перемножением двоек. Чтобы не запутаться, договорились обозначение в 1024 байт писать с большой буквы: Кб (сравни с км, кг).

209

Сколько разных слов длины 4 можно написать в двоичном алфавите? Построй дерево перебора, запиши ответ.

А сколько слов длины 5 можно написать в двоичном алфавите?

Во сколько раз увеличивается количество возможных слов при увеличении длины кода в двоичном алфавите на 1?

210

Для измерения информации используют килобайты, при этом $1 \text{ Кб} = 1024 \text{ байт}$. Сколько двоек нужно перемножить, чтобы получить число 1024? Сколько бит в 1 Кб? Сколько двоек надо перемножить, чтобы получить это число?

211

Вычисли, сколько байт в одном мегабайте, в одном гигабайте и в одном терабайте. 1 Кб больше 1000 байт на 24 байта.

Вычисли (см. на этой и следующей страницах):

На сколько 1 Мб больше, чем 1000 000 байт?

На сколько 1 Гб больше, чем 1000 000 000 байт?

На сколько 1 Тб больше, чем 1000 000 000 000 байт?
Поместится ли файл размером 1 Мб на флешку, на которую можно разместить миллион байт?

212



В 1963 году был принят стандарт двоичного кодирования ASCII (American Standard Code for Information Interchange), для кодирования каждого символа в этой кодировке использовался один байт. Произносится это сокращение как «аски», или «аски код».

В таблице кода были зафиксированы места для строчных и прописных букв латинского (то же, что английского) алфавита, цифр, знаков препинания и некоторых управляющих символов. Во второй части таблицы страны могли поставить буквы своего алфавита.

В начале 2000-х годов постепенно все перешли на другой стандарт кодирования Unicode — Юникод, в котором коды вдвое длиннее, поэтому этот код позволил зашифровать сразу почти все существующие алфавиты, иероглифы, математические символы и другие символы — теперь их можно использовать в одном файле. Сейчас наиболее распространенным является 32-битный вариант Юникода.

Используя таблицу кодов ASCII, запиши, расшифруй запись:

```
110010001110110111110100111011101111000011101100  
1110000011110010111010001110101011100000
```

Используя таблицу кодов ASCII, напиши двоичный код слова и предложения:

Информатика

Саша ответила: «My name is Sasha».

213

Современные компьютеры умеют работать не только с текстами, но и с изображениями, умеют записывать и воспроизводить музыку и видеофильмы. Попробуем разобраться, как кодируются изображения. Для кодирования чёрно-белой картинки придумали простые правила:

1. Картинка с помощью наложенной сетки разбивается на клетки.

2. Для каждой клетки решают, какой цвет ей присвоить: смотрят, больше в ней чёрного или белого цвета.

3. Теперь строят алгоритм построения картинку при помощи следующего двоичного кода:

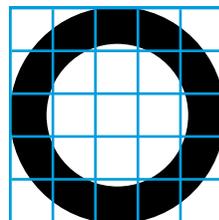
00 — встань в самую левую клетку сетки самого верхнего ещё не прочитанного ряда;

01 — закрась клетку белым и перейди к следующей клетке справа;

10 — закрась клетку чёрным и перейди к следующей клетке справа.

Пользуясь этими правилами, нарисуй картинку на клетчатой бумаге, выполнив следующий алгоритм:

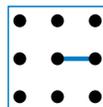
```
00 01 10 01 01 01 00 01 01 10
01 01 00 01 01 01 10 01 00 10
10 10 10 10 00 01 01 01 10 01
00 01 01 10 01 01 00 01 10 01
01 01
```



Закодируй картинку при помощи предложенного правила.

214

Найди выигрышную стратегию для Второго в игре *Ползунок* на поле 3×3 при условии, что Первый на первом ходу соединит центральную точку поля с одной из боковых.



215

Реши логическую задачу.

Жители двух соседних городов — города Рыцарей и города Лжецов — раз в год приезжают на ярмарку в город Хитрецов. Известно, что рыцари всегда говорят правду, лжецы всегда лгут, а хитрецы говорят правду через раз, т. е. из двух сказанных ими подряд предложений одно — правда, а другое — ложь. На ярмарке встретились трое из разных городов и затеяли спор:

Первый. *Один из нас рыцарь.*

Второй. *Да уж ты-то точно лжец.*

Третий. *Оба вы лжецы. Хотя я тоже не рыцарь.*

Определи, кто такие Первый, Второй и Третий, если точно известно, что все они живут в разных городах.

