

§ 1.3

Двоичное кодирование

Ключевые слова:

- дискретизация
- алфавит
- мощность алфавита
- двоичный алфавит
- двоичное кодирование
- разрядность двоичного кода

1.3.1. Преобразование информации из непрерывной формы в дискретную

Для решения своих задач человеку часто приходится преобразовывать имеющуюся информацию из одной формы представления в другую. Например, при чтении вслух происходит преобразование информации из дискретной (текстовой) формы в непрерывную (звук). Во время диктанта на уроке русского языка, наоборот, происходит преобразование информации из непрерывной формы (голос учителя) в дискретную (записи учеников).

Информация, представленная в дискретной форме, значительно проще для передачи, хранения или автоматической обработки. Поэтому в компьютерной технике большое внимание уделяется методам преобразования информации из непрерывной формы в дискретную.

Дискретизация информации — процесс преобразования информации из непрерывной формы представления в дискретную.

Рассмотрим суть процесса дискретизации информации на примере.



На метеорологических станциях имеются самопишущие приборы для непрерывной записи атмосферного давления. Результатом их работы являются кривые, показывающие, как изменялось давление в течение длительных промежутков времени (барограммы). Одна из таких кривых, вычерченная прибором в течение семи часов проведения наблюдений, показана на рис. 1.4.

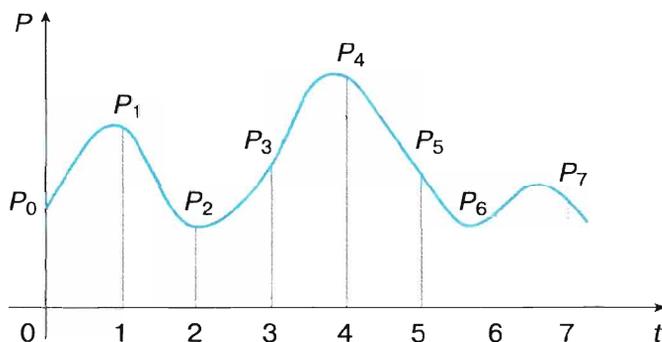


Рис. 1.4. Барограмма

На основании полученной информации можно построить таблицу, в которую будут занесены показания прибора в начале измерений и на конец каждого часа наблюдений (рис. 1.5).

t	0	1	2	3	4	5	6	7
P	P_0	P_1	P_2	P_3	P_4	P_5	P_6	P_7

Рис. 1.5. Таблица, построенная по барограмме

Полученная таблица даёт не совсем полную картину того, как изменялось давление за время наблюдений: например, не указано самое большое значение давления, имевшее место в течение четвёртого часа наблюдений. Но если занести в таблицу значения давления, наблюдаемые каждые полчаса или 15 минут, то новая таблица будет давать более полное представление о том, как изменялось давление.

Таким образом, информацию, представленную в непрерывной форме (барограмму, кривую), мы с некоторой потерей точности преобразовали в дискретную форму (таблицу).

В дальнейшем вы познакомитесь со способами дискретного представления звуковой и графической информации.

1.3.2. Двоичное кодирование

В общем случае, чтобы представить информацию в дискретной форме, её следует выразить с помощью символов какого-нибудь естественного или формального языка. Таких языков тысячи. Каждый язык имеет свой алфавит.

Алфавит — набор отличных друг от друга символов (знаков), используемых для представления информации. **Мощность алфавита** — это количество входящих в него символов (знаков).

Алфавит, содержащий два символа, называется **двоичным алфавитом** (рис. 1.6). Представление информации с помощью двоичного алфавита называют **двоичным кодированием**. Закодировав таким способом информацию, мы получим её **двоичный код**.

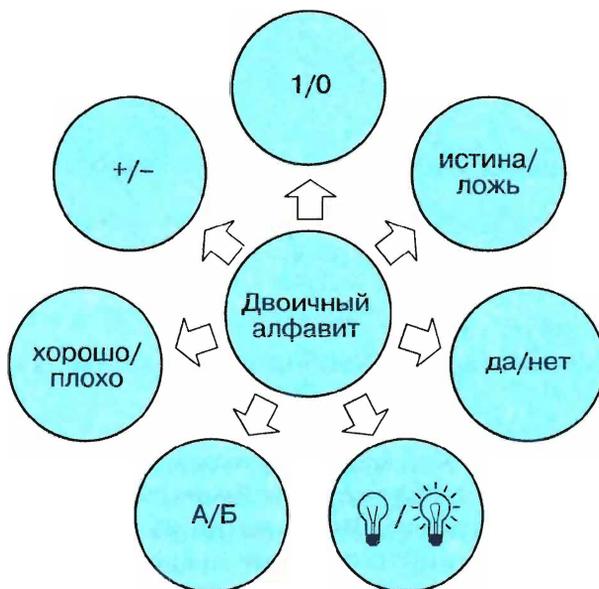


Рис. 1.6. Примеры символов двоичного алфавита

Рассмотрим в качестве символов двоичного алфавита цифры 0 и 1.

Покажем, что любой алфавит можно заменить двоичным алфавитом. Прежде всего, присвоим каждому символу рассматриваемого алфавита порядковый номер. Номер представим с помощью двоичного алфавита. Полученный двоичный код будем считать кодом исходного символа (рис. 1.7).



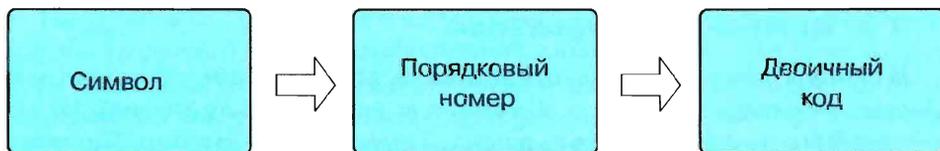


Рис. 1.7. Схема перевода символа произвольного алфавита в двоичный код

Если мощность исходного алфавита больше двух, то для кодирования символа этого алфавита потребуется не один, а несколько двоичных символов. Другими словами, порядковому номеру каждого символа исходного алфавита будет поставлена в соответствие цепочка (последовательность) из нескольких двоичных символов.

Правило двоичного кодирования символов алфавита мощности больше двух представим схемой на рис. 1.8.

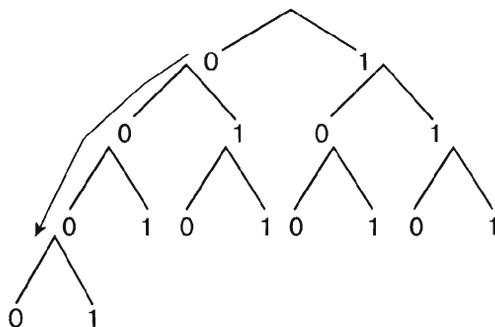


Рис. 1.8. Схематическое представление получения двоичных кодов

Двоичные символы (0, 1) здесь берутся в заданном алфавитном порядке и размещаются слева направо. Двоичные коды (цепочки символов) читаются сверху вниз. Все цепочки из двух двоичных символов (кодвые комбинации) позволяют представить четыре различных символа произвольного алфавита:

Порядковый номер символа	1	2	3	4
Двузначный двоичный код	00	01	10	11

Цепочки из трёх двоичных символов получаются дополнением двузначных двоичных кодов справа символом 0 или 1. В итоге

трёхзначных двоичных кодовых комбинаций получается 8 — вдвое больше, чем двузначных:

Порядковый номер символа	1	2	3	4	5	6	7	8
Трёхзначный двоичный код	000	001	010	011	100	101	110	111

Соответственно, четырёхзначный двоичный код позволяет получить 16 кодовых комбинаций, пятизначный — 32, шестизначный — 64 и т. д.

Длину двоичной цепочки — количество символов в двоичном коде — называют **разрядностью двоичного кода**.

Разрядность двоичного кода	1	2	3	4	5	6	7	8	9	10
Количество кодовых комбинаций	2	4	8	16	32	64	128	256	512	1024

Обратите внимание, что $2 = 2^1$, $4 = 2^2$, $8 = 2^3$, $16 = 2^4$, $32 = 2^5$ и т. д.

Если количество кодовых комбинаций обозначить буквой N , а разрядность двоичного кода — буквой i , то выявленная закономерность в общем виде будет записана так:

$$N = 2^i.$$

Задача. Вождь племени Мульти поручил своему министру разработать двоичный код и перевести в него всю важную информацию. Какой разрядности потребуется двоичный код, если алфавит, используемый племенем Мульти, содержит 16 символов? Выпишите все кодовые комбинации.

Решение. Так как алфавит племени Мульти состоит из 16 символов, то и кодовых комбинаций им нужно 16. В этом случае длина (разрядность) двоичного кода определяется из соотношения: $16 = 2^i$. Отсюда $i=4$.

Чтобы выписать все кодовые комбинации из четырёх 0 и 1, воспользуемся схемой на рис. 1.8: 0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111.



На сайте <http://school-collection.edu.ru/> размещена виртуальная лаборатория «Цифровые весы». С её помощью вы можете самостоятельно открыть метод разностей — один из способов получения двоичного кода целых десятичных чисел.

1.3.3. Универсальность двоичного кодирования

В начале этого параграфа вы узнали, что информация, представленная в непрерывной форме, может быть выражена с помощью символов некоторого естественного или формального языка. В свою очередь, символы произвольного алфавита могут быть преобразованы в двоичный код. Таким образом, с помощью двоичного кода может быть представлена любая информация на естественных и формальных языках, а также изображения и звуки (рис. 1.9). Это и означает универсальность двоичного кодирования.

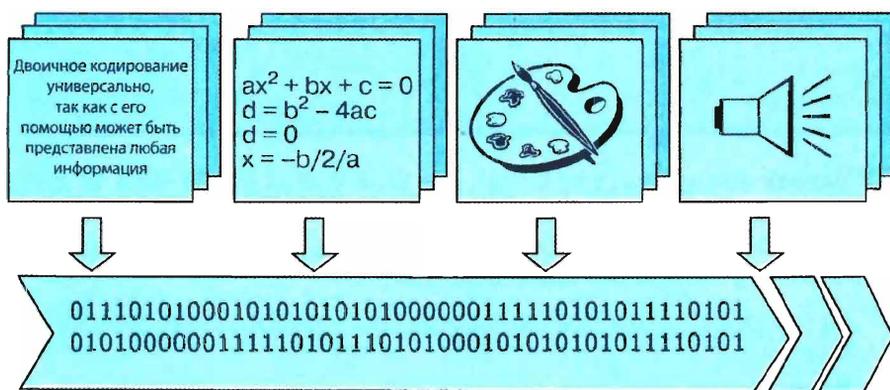


Рис. 1.9. Преобразование различных форм представления информации в двоичный код

Двоичные коды широко используются в компьютерной технике, требуя только двух состояний электронной схемы — «включено» (это соответствует цифре 1) и «выключено» (это соответствует цифре 0).

Простота технической реализации — главное достоинство двоичного кодирования. Недостаток двоичного кодирования — большая длина получаемого кода.

1.3.4. Равномерные и неравномерные коды

Различают равномерные и неравномерные коды. Равномерные коды в кодовых комбинациях содержат одинаковое число символов, неравномерные — разное.

Выше мы рассмотрели равномерные двоичные коды.

Примером неравномерного кода может служить азбука Морзе, в которой для каждой буквы и цифры определена последовательность коротких и длинных сигналов. Так, букве Е соответствует короткий сигнал («точка»), а букве Ш — четыре длинных сигнала (четыре «тире»). Неравномерное кодирование позволяет повысить скорость передачи сообщений за счёт того, что наиболее часто встречающиеся в передаваемой информации символы имеют самые короткие кодовые комбинации.

САМОЕ ГЛАВНОЕ

Дискретизация информации — процесс преобразования информации из непрерывной формы представления в дискретную. Чтобы представить информацию в дискретной форме, её следует выразить с помощью символов какого-нибудь естественного или формального языка.

Алфавит языка — набор отличных друг от друга символов, используемых для представления информации. **Мощность алфавита** — это количество входящих в него символов.

Алфавит, содержащий два символа, называется **двоичным алфавитом**. Представление информации с помощью двоичного алфавита называют двоичным кодированием. Двоичное кодирование универсально, так как с его помощью может быть представлена любая информация.

Вопросы и задания

1. С какой целью человек осуществляет преобразование информации из одной формы представления в другую? Приведите примеры таких преобразований.
2. В чём суть процесса дискретизации информации?
3. Что такое алфавит языка?
4. Что такое мощность алфавита? Может ли алфавит состоять из одного символа?



5. Какие символы могут входить в двоичный алфавит?
6. Сколько существует различных последовательностей из символов «плюс» и «минус» длиной ровно пять символов?
7. Почему двоичное кодирование является универсальным?
8. Как связаны мощность алфавита и разрядность двоичного кода, достаточного для кодирования всех символов этого алфавита?
9. Вождь племени Мульти поручил своему министру разработать двоичный код и перевести в него всю важную информацию. Достаточно ли пятиразрядного двоичного кода, если алфавит, используемый племенем Мульти, содержит 26 символов?
10. От разведчика была получена следующая шифрованная радиogramма, переданная с использованием азбуки Морзе:

- . . - . . - - . . - - - .

При передаче радиogramмы было потеряно разбиение на буквы, но известно, что в радиogramме использовались только следующие буквы:

И	А	Н	Г	Ч
..	.-	-.	--.	----.

Определите текст радиogramмы.