

ЛЕКЦИЯ 4

Кодирование информации. Системы счисления

Основное содержание:

- Кодирование информации. Системы счисления.

Кодирование информации — это процесс её представления в определённой форме для хранения, передачи и обработки, а системы счисления — это способы записи чисел с помощью набора символов (цифр).

Ключевые понятия: позиционная система счисления (значение цифры зависит от её положения, например, десятичная) и двоичный код (использование 0 и 1 для представления данных в компьютерах).

Цель кодирования — сделать информацию удобной для компьютерной обработки, что часто достигается переводом чисел из десятичной системы в двоичную.

Кодирование информации

Цель: Преобразовать информацию из одной формы в другую для удобства её использования, хранения или передачи.

Принцип: Использование определённой системы символов и правил для представления данных.

Примеры:

Для чисел: Перевод чисел из десятичной системы в двоичную для компьютерной обработки.

Для текста: Использование наборов символов (кодировок) для отображения букв, например, кириллицы.

• *Для изображений и звука:* Цифровое представление информации, удобное для электронных устройств.

Системы счисления

Определение: Набор символов (цифр) и правил их использования для записи чисел.

Позиционные системы счисления:

Значение каждой цифры зависит от её позиции в записи числа (например, в десятичной системе 100, где "1" на первом месте означает сотню).

Двоичная система счисления:

Символы: Использует всего две цифры — 0 и 1.

Применение: Является основой для представления информации в компьютерах, где "1" может означать наличие сигнала, а "0" — его отсутствие.

Преимущества: Простота реализации и передачи данных, несмотря на большую длину кода.

Как кодирование связано с системами счисления

- Системы счисления являются инструментом для кодирования числовой информации.
- Перевод чисел из одной системы счисления в другую (например, из десятичной в двоичную) — это процесс кодирования.
- В информатике числовая информация кодируется в двоичной системе для того, чтобы её могли обрабатывать компьютеры.

Информация редко используется человеком в чистом виде, она всегда как-то представлена – формализована или закодирована. Одна и та же информация может быть представлена в разных формах, а одни и те же символы нести разную смысловую нагрузку. В данной лекции мы рассмотрим различные типы счисления (в первую очередь - двоичную) и теоретические основы кодирования. Затем мы остановимся на сравнительном анализе различных методов кодирования, в том числе - методы сжатия цифровой информации и коды исправления ошибок.

1. СИСТЕМЫ СЧИСЛЕНИЯ Для машинной обработки информации её необходимо представить в какой-либо системе счисления. Система счисления – совокупность приёмов наименования и записи чисел с помощью цифр. Как мы обсудили в первой лекции, счётные устройства начали создаваться задолго до возникновения и алгебры логики, и теории алгоритмов. И определяющую роль в их конструкции играли выбранные для них системы счисления.

Первые механические счётные машины (суммирующая машина Паскаля (1642), счётная машина Лейбница (1673), аналитическая машина Бэббиджа (1848)), были разработаны на основе десятичной системы счисления. Эти механические системы были громоздки. Так, если бы проект машины Бэббиджа был реализован, то по размерам машина сравнялась бы с локомотивом, и чтобы привести в движение её устройства, понадобился бы паровой двигатель. Причинами этого были механический принцип построения устройств и использование десятичной системы счисления, затрудняющей создание простой элементной базы. Через 63 года после смерти Бэббиджа немецкий студент Конрад Цузе начал работу по созданию машины, основанной на принципах действия Аналитической машины Бэббиджа. В 1937 машина Z1 была готова. Работала она на основе двоичной системы счисления и была чисто механической. Но использование двоичной системы сильно уменьшило размеры устройства – машина занимала всего два квадратных метра.

В современных компьютерах вся информация также хранится в виде последовательностей нулей и единиц. Однако двоичная система счисления в чистом виде обладает рядом принципиальных недостатков, тормозящих развитие компьютерной техники. Главными из этих недостатков являются проблема представления отрицательных чисел и «нулевая избыточность» (т.е. отсутствие избыточности, из чего вытекает невозможность определения, произошло ли искажение информации при её передаче). Практическая потребность в решении этих вопросов вызывает сегодня повышенный интерес к способам представления информации в компьютере и новым компьютерным арифметикам. Так, например, Джон фон Нейман доказал теорему о том, что троичная система счисления позволяет наиболее эффективно среди всех основных позиционных систем счисления

«сворачивать» информацию о вещественном числе. Ниже мы рассмотрим основные системы счисления, используемые в компьютере.

1.1 Позиционные системы счисления Система счисления или нумерация – способ записи (обозначения) чисел. Система счисления – символический метод записи чисел, представление чисел с помощью письменных знаков. Система счисления: даёт представления множества чисел (целых или вещественных); даёт каждому числу уникальное представление (или, по крайней мере, стандартное представление); отражает алгебраическую и арифметическую структуру чисел. Символы, при помощи которых записываются числа, называются цифрами, а их совокупность – алфавитом системы счисления. Количество цифр, составляющих алфавит, называется его размерностью. Система счисления называется позиционной, если количественный эквивалент цифры зависит от её положения в записи числа.

Позиционная система счисления – система счисления, в которой один и тот же числовой знак (цифра) в записи числа имеет различные значения в зависимости от того места (разряда), где он расположен. Число – абстракция, используемая для количественной характеристики объектов. Возникнув ещё в первобытном обществе из потребностей счёта, понятие числа изменялось и обогащалось и превратилось в важнейшее математическое понятие.

Цифры – знаки для записи чисел (числовые знаки). Слово «цифра» без уточнения обычно означает один из следующих знаков: 0 1 2 3 4 5 6 7 8 9 (т. н. «арабские цифры»). Существуют также много других вариантов: римские цифры (I V X L C D M), шестнадцатеричные цифры (0 1 2 3 4 5 6 7 8 9 A B C D E F), в некоторых языках, например, в древнегреческом, в иврите, в церковнославянском, существует система записи чисел буквами и др. Виды систем счисления: позиционные (значение цифры зависит от её позиции в изображении числа): двоичная, троичная,..., десятичная,..., шестнадцатеричная,..., шестидесятиричная и т.д.; непозиционные (значение цифры не зависит от места, занимаемого в изображении числа: единичная, римская, алфавитные).

Пример 1: Римская система счисления: V X L 5 10 C D M 50 100 500 1000 XXVIII=10+10+5+1+1+1=28; XCIX= -10+100-1+10=99 В привычной нам десятичной системе значение числа образуется следующим образом: значения цифр умножаются на «веса» соответствующих разрядов и все полученные значения складываются.

Например, $5047 = 5 \cdot 1000 + 0 \cdot 100 + 4 \cdot 10 + 7 \cdot 1$. Такой способ образования значения числа называется аддитивно мультипликативным. Последовательность чисел, каждое из которых задаёт «вес» соответствующего разряда, называется базисом позиционной системы счисления. Основное достоинство позиционной системы – возможность записи произвольного числа при помощи ограниченного количества символов. Позиционную систему счисления называют традиционной, если её базис образуют члены геометрической прогрессии, а значения цифр есть целые неотрицательные числа. Позиционная система: число x может быть представлено в системе с основанием p , как $x = a_n \cdot p^n + a_{n-1} \cdot p^{n-1} + \dots + a_1 \cdot p^1 + a_0 \cdot p^0$, где $a_n, \dots, a_0$ – цифры в представлении данного числа, в основание системы счисления.

Пример 2. $103510 = 1 \cdot 103 + 0 \cdot 102 + 3 \cdot 101 + 5 \cdot 100$ Так, базисы десятичной, двоичной и восьмеричной систем счисления образуют геометрические прогрессии со знаменателями 10, 2 и 8 соответственно. В общем виде базис традиционной системы счисления можно записать так: ...P-3, P-2, P-1, 1, P, P2, P3,..., Pn,... (1)

Знаменатель P геометрической прогрессии, члены которой образуют базис традиционной системы счисления, называются основанием этой системы счисления. Традиционные системы счисления с основанием P иначе называют P -ичными.

Десятичная	Двоичная	Восьмеричная	Шестнадцатеричная
0	0	0	0
1	1	1	1
2	10	2	2
...	...	...	...
10	1010	12	A
11	1011	13	B
12	1100	14	C
13	1101	15	D
14	1110	16	E
15	1111	17	F
16	10000	20	10
...	...	...	...
1000	1111101000	1750	3E8

В P -ичных системах размерность алфавита равна основанию системы счисления. Так, алфавит десятичной системы составляют цифры 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Алфавитом произвольной системы счисления с основанием P служат числа 0, 1, ..., $P-1$, каждое из которых должно быть записано с помощью одного уникального символа, младшей цифрой всегда является 0. В класс позиционных систем счисления входят также системы, в которых либо базис не является геометрической прогрессией, а цифры есть целые неотрицательные числа, либо базис является геометрической прогрессией, но цифры не являются целыми неотрицательными числами. К первым можно отнести факториальную и фибоначчьеву системы счисления, ко вторым – уравновешенные системы счисления. Такие системы будем называть нетрадиционными. Алфавитом фибоначчевой системы являются цифры 0 и 1, а её базисом – последовательность чисел Фибоначчи 1, 2, 3, 5, 8, 13, 21, 34, 55, 89... Базисом факториальной системы счисления является последовательность $1!, 2!, \dots, n!, \dots$. Здесь количество цифр, используемых в разряде, увеличивается с ростом номера разряда. Любое десятичное число можно представить в любой позиционной системе счисления, а для целых чисел в большинстве систем это можно сделать единственным образом. Любое натуральное число можно записать в какой угодно P -ичной системе счисления, причём единственным образом. Рассмотрим запись чисел в разных системах счисления. Двоичная (используются цифры 0, 1); восьмеричная (используются цифры 0, 1, ..., 7); шестнадцатеричная (для первых целых чисел от нуля до девяти используются цифры 0, 1, ..., 9, а для следующих чисел – от десяти до пятнадцати – в качестве цифр используются символы A, B, C, D, E, F).

Существуют два основных формата представления чисел в памяти компьютера: для кодирования целых чисел и для кодирования действительных чисел

$$A_q = \pm \sum a_i q^i \quad (2)$$

$$A_q = \pm (a_{n-1} q^{n-1} + a_{n-2} q^{n-2} + \dots + a_0 q^0 + a_{-1} q^{-1} + \dots + a_{-m} q^{-m}) \quad (3)$$

Здесь A – само число, q – основание системы счисления, a_i – цифры, принадлежащие алфавиту данной системы счисления, n – число целых разрядов числа, m – число дробных разрядов числа. $M = a_{n-1} a_{n-2} a_{n-3} \dots a_3 a_2 a_1 a_0$ – запись числа $a_{n-1} q^{n-1} + a_{n-2} q^{n-2} + \dots + a_1 q^1 + a_0 q^0 + a_{-1} q^{-1} + \dots + a_{-m} q^{-m}$, где a_i – цифры системы счисления, n и m – число целых и дробных разрядов, соответственно. Пример 3. $1011_2 = 1 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 = 1 \cdot 2 + 1 \cdot 2^{-1}$; $276,528_{10} = 2 \cdot 8^2 + 7 \cdot 8^1 + 6 \cdot 8^0 + 5 \cdot 8^{-1} + 2 \cdot 8^{-2}$; $M = 342$ $q = 10$ $M = 3 \cdot 10^2 + 4 \cdot 10^1 + 2 \cdot 10^0$ $q = 5$ $M = 3 \cdot 5^2 + 4 \cdot 5^1 + 2 \cdot 5^0$ (9710) $q = 8$ $M = 3 \cdot 8^2 + 4 \cdot 8^1 + 2 \cdot 8^0$ (22610)

1	=	1 ₂ ;
2	=	10 ₂
3	=	11 ₂
4	=	100 ₂
5	=	101 ₂
6	=	110 ₂
7	=	111 ₂
8	=	1000 ₂
9	=	1001 ₂
10	=	1010 ₂
11	=	1011 ₂
12	=	1100 ₂
13	=	1101 ₂
14	=	1110 ₂
15	=	1111 ₂
16	=	10000 ₂

Рис. 1.

Первые 16 чисел, представленные в двоичной системе. В десятичной системе счисления обыкновенная дробь будет точно представима конечной дробью, если существует такое натуральное число m , при умножении на которое знаменателя дроби можно получить некоторую натуральную степень числа 10. Если же такого числа не существует, то эту дробь можно представить только в виде бесконечной периодической дроби.

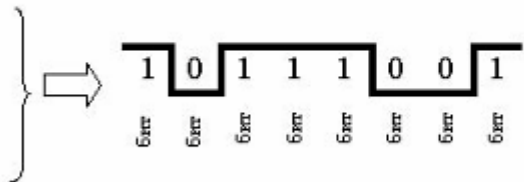
Пример 4. Известно, что правильную десятичную дробь $1/3$ нельзя записать в виде конечной десятичной дроби. Однако в троичной и 9-ричной системах счисления эта дробь будет записана в виде конечной Р-ичной дроби. В 3-ичной системе счисления: $1/3 = 0,13$; в 9-ричной: $1/3 = 0,39$.

Десятичная обыкновенная дробь будет точно представима конечной Р-ичной дробью, если существует такое натуральное число m , при умножении на которое знаменателя дроби можно получить некоторую натуральную степень числа P . Если же такого числа не существует, то в Р-ичной системе счисления дробь окажется бесконечной периодической.

Пример 5. Запишем $5/16$ в двоичной системе. В знаменателе уже стоит четвёртая степень двойки. Переведём числитель в двоичную систему ($5=101_2$) и дополним получившееся число до четырёх цифр: 0101 . В результате получим: $5/16=0,0101_2$.

Пример 6. Запишем $0,011_2 = \frac{11_2}{1000_2}$ в десятичной системе. Для того, чтобы знаменатель, равный сейчас 23, оказался степенью десяти, его нужно домножить на 5^3 , тогда

$$\frac{11_2}{1000_2} = \frac{3}{8} = \frac{3 \cdot 5^3}{8 \cdot 5^3} = \frac{375}{1000} = 0,375$$

Виды информации	Двоичный код
Числовая	
Текстовая	
Графическая	
Звуковая	
Видео	

Двоичная система счисления - позиционная система счисления с основанием 2. В этой системе 2 счисления натуральные числа записываются с помощью всего лишь двух символов (в роли которых обычно выступают цифры 0 и 1). В этой лекции нас в основном будет интересовать двоичная система счисления, точнее - двоичный (цифровой) код, основанный на двоичной системе счисления, использующий для представления буквенно-цифровых и других символов наборы комбинаций цифр 1 и 0. В двоичной системе счисления всего две цифры, называемые двоичными. Название разряда двоичного числа – бит. Веса разрядов в двоичной системе изменяются по степеням двойки. Поскольку вес каждого разряда уменьшается либо на 0, либо на 1, то в результате значение числа определяется как сумма соответствующих значений степеней двойки. $1010_2 = 1 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0 = 10$. Известны следующие преимущества использования двоичной системы в вычислительной технике: простота технической реализации; надёжность и помехоустойчивость; возможность применения аппарата булевой алгебры для выполнения логических операций; простота правил двоичной арифметики.

Двоичная система используется в цифровых устройствах, т.к. является наиболее простой и соответствует требованиям:

1) Чем меньше значений существует в системе, тем проще изготовить отдельные элементы, оперирующие этими значениями. В частности, две цифры двоичной системы счисления могут быть легко представлены многими физическими явлениями: есть ток — нет тока, индукция магнитного поля больше пороговой величины или нет и т. д.

2) Чем меньше количество состояний у элемента, тем выше помехоустойчивость и тем быстрее он может работать. Например, чтобы закодировать три состояния через величину индукции магнитного поля, потребуется ввести два пороговых значения, что не будет способствовать помехоустойчивости и надёжности хранения информации.

3) Двоичная арифметика является довольно простой. Простыми являются таблицы сложения и умножения - основных действий над числами B и m .

4) возможно применение аппарата алгебры логики для выполнения побитовых операций над числами.