

ЛЕКЦИЯ

Тема: Основные способы и методы кодирования информации

Дисциплина: Основы кодирования

ОП Системы информационной безопасности

Авторы: Асоц. проф. КиИИ Даненова Г.Т.
Асоц. проф. КиИИ Коккоз М.М.
Асоц. проф. КиИИ Кан О.А.
Асоц. проф. СМиТ Ахметжанов Т.Б.

План лекции

1. Алфавитное кодирование информации.
2. Равномерное кодирование.
3. Неравномерное кодирование.
4. Префиксные коды.
5. Двоично-десятичные коды.
6. Арифметические действия над двоично-десятичными числами.
7. Среда разработки Visual Studio.Net.

Алфавитное кодирование информации

Алфавитное кодирование – это перевод информации с одного языка на другой.

В вычислительной технике рассматривается перевод информации, представленной в виде набора символов определенного языка в двоичный код. Например, перевод символов английского языка в двоичный код.

При кодировании набора символов, в двоичный код каждый символ заменяется набором нулей и единиц (двоичной комбинацией).

3

В общем случае задачу алфавитного кодирования можно представить следующим образом. Пусть заданы два **алфавита** **A** и **B**, состоящие из конечного числа символов:
 $A = \{a_1, a_2, \dots, a_n\}$ и $B = \{b_1, b_2, \dots, b_k\}$.

Элементы алфавита называются **буквами**. Упорядоченный набор букв в алфавите называется **словом**.

Количество букв в слове называется **длиной** слова.

4

Базовая таблица ASCII кодов

Базовая таблица кодировки ASCII

32 пробел	48 0	64 @	80 P	96 `	112 p
33 !	49 1	65 A	81 Q	97 a	113 q
34 "	50 2	66 B	82 R	98 b	114 r
35 #	51 3	67 C	83 S	99 c	115 s
36 \$	52 4	68 D	84 T	100 d	116 t
37 %	53 5	69 E	85 U	101 e	117 u
38 &	54 6	70 F	86 V	102 f	118 v
39 '	55 7	71 G	87 W	103 g	119 w
40 (56 8	72 H	88 X	104 h	120 x
41)	57 9	73 I	89 Y	105 i	121 y
42 *	58 :	74 J	90 Z	106 j	122 z
43 +	59 ;	75 K	91 [107 k	123 {
44 ,	60 <	76 L	92 \	108 l	124
45 -	61 =	77 M	93]	109 m	125 }
46 .	62 >	78 N	94 ^	110 n	126 ~
47 /	63 ?	79 O	95 _	111 o	127

7

Первая буква слова называется **началом**, или **префиксом**, а последняя буква слова — **окончанием**, или **постфиксом**.

Суть кодирования заключается в том, что каждой букве алфавита **A** сопоставляется слово из алфавита **B** согласно правилу кодирования.

Пример: Представить английскую букву A двоичным кодом. Используя таблицу ASCII кодов, получим $A = 1000001$.

Алфавитное кодирование бывает равномерным и неравномерным.

5

Равномерное кодирование

Код называется **равномерным** (или кодом **постоянной длины**), если все его слова содержат одинаковое число букв.

Например: ASCII-код, все символы закодированы с помощью одного байта.

Об эффективности равномерного кода можно говорить только в том случае, когда вероятности появления символов источника равны, то есть его энтропия максимальна.

6

Например, если выбрать в качестве алфавита B двоичный код, то трехразрядным кодом можно закодировать 8 символов, а восьмиразрядным кодом – 256 символов.

Первым равномерным алфавитным кодом является телеграфный код Бодо. Каждая буква записывалась пятью битами.

Следовательно, кодом Бодо можно было передать $2^5=32$ различных символов.

9

Благодаря идее Бодо удалось автоматизировать процесс передачи и печати букв. Был создан клавишный телеграфный аппарат. Нажатие клавиши с определенной буквой вырабатывает соответствующие сигналы из пяти импульсов, которые передаются по линии связи. Принимающий аппарат под воздействием этих сигналов печатает ту же букву на бумажной ленте.

Код Морзе

А • —	Л • — • •	Ц — • — •
Б — • • • •	М — —	Ч — — — •
В • — —	Н — •	Ш — — — —
Г — — •	О — — —	Щ — — • —
Д — • •	П • — — •	Ъ • — — • — •
Е •	Р • — •	Ы — • — —
Ж • • • —	С • • • •	Ь — • • —
З — — • •	Т —	Э • • — • •
И • •	У • • —	Ю • • — —
Й • — — —	Ф • • — •	Я • — • —
К — • —	Х • • • •	

В 1938 году Сэмюэл Морзе изобрел релейное устройство и кодовую азбуку. Принцип кодирования азбуки Морзе исходит из того, что буквы, которые чаще употребляются в английском языке, кодируются более короткими сочетаниями точек и тире. Это делает передачи эффективнее. Точки и тире в коде Морзе передаются с помощью радиосигналов или прерыванием постоянного электрического тока.



15

Чтобы избежать неоднозначности кода, при кодировании нужно следовать **условию Фано**, которое гласит: никакое кодовое слово не должно являться началом другого кодового слова. Так в предыдущем примере код символа "а" = 1 является началом кода символа "с" = 11. Для исключения ошибок декодирования применяют разделимые схемы.

17

Разделимые схемы

Схема называется разделимой, если любое слово состоящее из элементарных кодов символов, единственным образом разлагается на элементарные коды. Алфавитное кодирование с разделимой схемой дает правильное, однозначное декодирование.

Если таблица кодов не удовлетворяет условию Фано, то схема заведомо не является разделимой.

18

Префиксные коды

Важным классом однозначно декодируемых неравномерных кодов являются префиксные коды.

Префиксные коды - это такие алфавитные коды, где ни один элементарный код не является началом (префиксом) другого элементарного кода.

Например, $a = 0, b = 01$. Это не префиксный код.

$a = 01, b = 101$. Это префиксный код.

21

Преимущества префиксных кодов

1. Однозначно декодируемый и делимый код.
2. Удастся получить более короткие коды, чем с помощью равномерного кода фиксированной длины.
3. Возможности декодирования сообщения, не получая его целиком, а по мере его поступления.

25

Арифметические действия над двоично-десятичными числами

Поскольку человеку наиболее привычны представление и арифметические действия в десятичной системе счисления, а для компьютера двоичное представление и двоичная арифметика, то была введена промежуточная система двоично-десятичной записи чисел.

Такая система чаще всего применяется там, где существует необходимость частого использования процедуры десятичного ввода-вывода (электронные часы, калькуляторы, АОНы, и т.д.).

39

Рассмотрим пример сложения двух чисел в двоично-десятичном коде.

Десятичные	Двоично-десятичные
7 9 4 6	0111 1001 0100 0110
+ 1 8 5 8	0001 1000 0101 1000
-----	-----
8 17 9 14	01000 10001 01001 01110
Сумма без переносов	
9 8 0 4	1001<1000<0000<0100
Сумма с переносами	

41

Преимущества двоично-десятичной системы счисления

Преобразование чисел из десятичной системы в двоично-десятичную систему счисления не связано с вычислениями и его легко реализовать, используя при этом простейшие электронные схемы, так как преобразовывается небольшое количество (4) двоичных цифр.

Обратное преобразование происходит в ЭВМ автоматически с помощью простой программы перевода.

43

Контрольные вопросы

1. Что такое алфавитное кодирование?
2. В чем отличие равномерного кодирования от неравномерного?
3. Что такое префиксные коды?
4. Как строятся двоично-десятичные коды?
5. Назовите преимущества двоично-десятичных кодов.
6. В чем особенность кода Грея?

Следующая лекция

Корректирующие и систематические коды кодирования информации