常用校验码（奇偶校验码、海明校验码、CRC 校验码）

计算机系统运行时各个部之间要进行数据交换。为确保数据在传送过程正确无误常使用检验码。我们常使用的检验码有三种，分别是**奇偶校验码**、**海明校验码**和**循环冗余校验码(CRC)**。

奇偶校验码

奇偶校验码最简单，但只能检测出奇数位出错。如果发生偶数位错误就无法检测。但经研究是奇数位发生错误的概率大很多。而且奇偶校验码无法检测出哪位出错，所以属于无法矫正错误的校验码。奇偶校验码是奇校验码和偶校验码的统称。它们都是通过在要校验的编码上加一位校验位组成。如果是奇校验加上校验位后编码中 1 的个数为奇数个。如果是偶校验加上校验位后编码中 1 的个数为偶数个。

例：

原编码	奇校验	偶校验
0000	0000 1	0000 0
0010	0010 0	0010 1
1100	1100 1	1100 0
1010	1010 1	1010 0

如果发生奇数个位传输出错那么编码中 1 的个数就会发生变化。从而校验出错误，要求从新传输数据。目前应用的奇偶校验码有 3 种：

水平奇偶校验码对每一个数据的编码添加校验位，使信息位与校验位处于同一行（如上例）。

垂直奇偶校验码把数据分成若干组，一组数据排成一行再加一行校验码。针对每一行列采用奇校验或偶校验。

例：有 32 位数据 **10100101 00110110 11001100 10101011**

	垂直奇校验	垂直偶校验
数据	**10100101**	**10100101**
	00110110	**00110110**
	11001100	**11001100**
	10101011	**10101011**
校验	00001011	11110100

水平垂直奇偶校验码就是同时用水平校验和垂直校验。

例：

	奇校验	奇水平	偶校验	偶水平
数据	**10100101**	1	**10100101**	0
	00110110	1	**00110110**	0
	11001100	1	**11001100**	0
	10101011	0	**10101011**	1
校验	00001011	0	11110100	1

海明校验码

海明码也是利用奇偶性来校验数据的。它是一种多重奇偶校验检错系统，它通过在数据位之间插入 k 个校验位来扩大码距从而实现检错和纠错。

设原来数据有 n 位，要加入 k 位校验码。怎么确定 k 的大小呢？ k 个校验位可以有 pow(2,k) (代表 2 的 k 次方) 个编码，其中有一个代表是否出错，剩下 pow(2,k)-1 个编码则用来表示到底是哪一位出错。因为 n 个数据位和 k 个校验位都可能出错，所以 k 满足 pow(2,k)-1 >= n+k。

设 k 个校验码为 P1,P2...Pk 。n 个数据位为 D0,D1...Dn 。产生的海明码为 H1,H2...H(n+k) 。如有 8 个数据位，根据 pow(2,k)-1 >= n+k 可以知道 k 最小是 4。那么得到的海明码是：

H12 H11 H10 H9 H8 H7 H6 H5 H4 H3 H2 H1
D7 D6 D5 D4 P4 D3 D2 D1 P3 D0 P2 P1

然后怎么知道 Pi 校验哪个位呢？自己可以列个校验关系表：

海明码	下标	校验位组
H1(P1)	1	P1
H2(P2)	2	P2
H3(D0)	1+2	P1,P2
H4(P3)	4	P3
H5(D1)	1+4	P1,P2
H6(D2)	2+4	P2,P3
H7(D3)	1+2+4	P1,P2,P3
H8(P4)	8	P4
H9(D4)	1+8	P1,P4
H10(D5)	2+8	P2,P4
H11(D6)	1+2+8	P1,P2,P4
H12(D7)	4+8	P3,P4

从表中可以看出：

P1 校验 P1，D0，D1，D3，D4，D6

P2 校验 P2，D0，D1，D2，D3，D5，D6

P3 校验 P3，D2，D3，D7

P4 校验 P4，D4，D5，D6，D7

其实上表很有规律很容易记，要知道海明码 Hi 由哪些校验组校验，可以把 i 化成二进制数，数中哪些位 k 是 1 就有哪些 Pk 校验。

如：H7 7=0111 所以由 P1,P2,P3。H11 11=1011 所以由 P1,P2,P4。 H3 3=0011 所以由 P1,P2。

那看看 Pi 的值怎么确定，如果使用偶校验，则：

P1=D0 xor D1 xor D3 xor D4 xor D6

P2=D0 xor D1 xor D2 xor D3 xor D5 xor D6

P3=D1 xor D2 xor D3 xor D7

P4=D4 xor D5 xor D6 xor D7

其中 xor 是异或运算，奇校验的话把偶校验的值取反即可。那怎么校验错误呢？其实也很简单，先做下面运算：

G1 = P1 xor D0 xor D1 xor D3 xor D4 xor D6

G2 = P2 xor D0 xor D1 xor D2 xor D3 xor D5 xor D6

G3 = P3 xor D1 xor D2 xor D3 xor D7

G4 = P4 xor D4 xor D5 xor D6 xor D7

例：已知被校验的数据为 6 位二进制数，D=101101，求其海明码表示方法。

若有 k 位数据，需要 n 位校验位 满足关系 k+n<=2^n，此处 k=6，则 n=4

海明码是由数据与校验位组合而成的。其组合规则为：将数据与校验码自左至右进行编码（$D_1D_2D_3...D_{10}$，下标是编号），其中编号为 2 的幂的位均为校验位(第 1、2、4、8 位)，其余为数据位。

则生成的码字为：

D_1 D_2 D_3 D_4 D_5 D_6 D_7 D_8 D_9 D_{10} （注：k+n=6+4=10，所以有 10 位编码）

a b 1 c 0 1 1 d 0 1 （注：a，b，c，d 为加入的检验码）

再将每一数据位的编号展开成 2 的幂的和：

3=2+1； 5=4+1； 6=4+2； 7=4+2+1； 9=8+1； 10=8+2

可以这样理解：编号为 3 的数据位与编号为 1 和 2 的校验位有关。

然后逆向来关注校验位：

a(也即编号为 1 的校验位)与编号为 3,5,7,9 的数据位有关,则 a=d3^d5^d7^d9(^是异或运算的意思)=1^0^1^0=0

（* 异或运算法则：两数相同为假，不同为真。0 代表假，1 代表真。0^1=1，0^0=0，1^1=0，1^0=1）

同理：

b(也即编号为 2 的校验位)与编号为 6，7，10 的数据位有关，则 a=d6^d7^d10=1^1^1=1

c(也即编号为 4 的校验位)=d5^d6^d7=0^1^1=0

d(也即编号为 8 的校验位)=d9^d10=0^1=1

求出 a，b，c，d 后就可以得到编好的海明码了，结果是：0110011101

循环冗余校验码

CRC 码利用生成多项式为 k 个数据位产生 r 个校验位进行编码，其编码长度为 n=k+r 所以又称(n，k)码。CRC 码广泛应用于数据通信领域和磁介质存储系统中。现在简单介绍下它的原理：

在 k 位信息码后接 r 位校验码，对于一个给定的(n，k)码。可以证明(数学高手自己琢磨证明过程)存在一个最高次幂为 n-k=r 的多项式 g(x)，根据 g(x)可以生成 k 位信息的校验码，g(x)被称为生成多项式。

用 C(x)=C(k-1)C(k-2)...C0 表示 k 个信息位，把 C(x)左移 r 位，就是相当于 C(x)*pow(2，r)给校验位空出 r 个位来了。给定一个生成多项式 g(x)，可以求出一个校验位表达式 r(x)。C(x)*pow(2，r) / g(x) = q(x) + r(x)/g(x) 用 C(x)*pow(2，r)去除生成多项式 g(x)商为 q(x)余数是 r(x)。所以有 C(x)*pow(2，r) = q(x)*g(x) + r(x)

C(x)*pow(2，r) + r(x)就是所求的 n 位 CRC 码，由上式可以看出它是生成多项式 g(x)的倍式。所以如果用得到的 n 位 CRC 码去除 g(x)如果余数是 0，就证明数据正确。否则可以根据余数知道出错位。

在 CRC 运算过程中，四则运算采用 mod2 运算(后面介绍)，即不考虑进位和借位。所以上式等价于 C(x)*pow(2，r) + r(x) = q(x)*g(x)

继续前先说下基本概念吧。

1.多项式和二进制编码

x 的最高次幂位对应二进制数的最高位。以下各位对应多项式的各幂次。有此幂次项为 1，无为 0。x 的最高幂次为 r 时，对应的二进制数有 r+1 位，例如：g(x)=pow(x，4) + pow(x，3) + x + 1 对应二进制编码是 11011。

2.生成多项式

是发送方和接受方的一个约定，也是一个二进制数。在整个传输过程中这个数不会变。

在发送方利用生成多项式对信息多项式做模 2 运算生成校验码。

在接受方利用生成多项式对收到的编码多项式做模 2 运算校验和纠错。

生成多项式应满足：

a.生成多项式的最高位和最低位必须为 1

b.当信息任何一位发生错误时被生成多项式模 2 运算后应该使余数不为 0

c.不同位发生错误时应该使余数不同

d.对余数继续做模 2 除应使余数循环

生成多项式很复杂，不过不用我们生成。

下面给出一些常用的生成多项式表

n k 二进制码(自己根据多项式和二进制编码的介绍)

7 4 1011 或 1101

7 3 11011 或 10111

15 11 1011

31 26 100101

3.模 2 运算

a.加减法法则

0 +/- 0 = 0

0 +/- 1 = 1

1 +/- 0 = 1

1 +/- 1 = 0

注意:没有进位和借位

b.乘法法则

利用模 2 加求部分积之和没有进位

c.除法法则

利用模 2 减求部分余数，没有借位，每商 1 位则部分余数减 1 位，余数最高位是 1 就商 1，不是就商 0，当部分余数的位数小于余数时该余数就是最后余数。

例：1110

1011)1100000

1011

1110

1011

1010

1011

0010(每商 1 位则部分余数减 1 位，所以前两个 0 写出)

0000

010(当部分余数的位数小于余数时，该余数就是最后余数)

最后商是 1110 余数是 010

下面讲下 CRC 的实际应用。

例：给定的生成多项式 g(x)=1011，用(7，4)CRC 码对 C(x)=1010 进行编码。

由题目可以知道下列的信息：

C(x)=1010，n=7，k=4，r=3，g(x)=1011 C(x)*pow(2，3)=1010000 C(x)*pow(2，3) / g(x) = 1001 + 11/1011 所以 r(x)=011。所以要求的编码为 1010011。

例 2：上题中数据传输后变为 1000011。试用纠错机制纠错 1000011 / g(x) = 1011 + 110/1011

不能整除所以出错了。因为余数是 110 查 1011 出错位表可以知道是第 5 位出错，对其求反即可。

例 3：已知要传送的数据是 859D，生成多项式是 10011B，求 CRC 校验码；实际传送的码序列是什么样的？
首先你应该知道一件事情，859D 这里的”D”表示什么？生成多项式是 10011B 的”B”又表示什么？
下面我来解释一下：
”D”表示十进制
”B”表示二进制
”O”表示八进制
”H”表示十六进制
好了，这个我们知道了，那么下一步我们就是要把十进制的 859 化成二进制（859D）
859D 化成二进制为：1101011011
OK，生成多项式是二进制，我们就不用化了，是五位（10011）
如果还要求 CRC 校验码；实际传送的码序列的话：
我们就在 1101011011 后面加 4 位，比刚才生成多项式少 1 位
于是就成了 11010110110000 再去除于生成多项式(10011)
求余数．余数为：1110 (二进制数相除相当于除数和被除数异或运算)
CRC 校验码 1110
实际传送的码序列 11010110111110