

# ISBN 校验的运行机制及其局限

余训培

(北京大学 信息管理系, 北京 100871)

**摘 要:**简要地介绍了 ISBN 校验的背景, 解释了校验在保证录入数据准确性方面的运行机制及其作用, 同时指出了由加和与权值导致的 ISBN 校验出错的情况; 提出应对 ISBN 校验的局限性加以重视。

**关键词:**ISBN 校验; 代码; 权值; 程序

**中图分类号:**G91      **文献标识码:**A

## 1 背景介绍

所谓校验, 就是在原代码后面加上一个校验码。校验代码是根据给定的原代码, 通过预先确定的包含了一定数学算法的计算机程序给出的, 然后附加在原代码的右端, 在编目时, 把这一完整的代码输入计算机后, 会由计算机根据算法确定出校验码, 然后和给定的校验码作比较, 如果二者一致, 说明输入的代码没有问题, 如果有出入, 说明中间过程出了差错, 需要进行进一步的核对。设置校验码是为了保证录入数据正确性。

ISBN(International Standard Book Number), 国际标准书号, 号码为 10 位, 由四段构成, 段间由连字符分隔, 如《图书馆自动基础教程》的 ISBN 号 7-301-03210-2。第一段为地域编号, 指示按语言或地理划分的地理区域; 第二段为出版社编号; 第三段为图书编号; 第四段为校验码(占一位)。除了校验码这一位, 其余九位代码可任意分配到其他三个段中, 表示地区、出版社和图书种类的灵活变化。我国的图书基本上使用 ISBN 校验方法。

## 2 ISBN 校验方法介绍

号码的校验位的确定是由前九位代码值算出来的, 用  $N(i)$  表示第  $i$  个代码的值, 给其指定权值  $11-i$ , 对前九位代码加权求和:  $S = \sum (11-i)N(i)$ , 其中  $i$  从 1 到 9 变化, 则校验码  $C = 11 - S \bmod 11$ 。当  $C = 10$  时, 取字符“X”为校验码, 否则, 指定相应的数字为校验码。如《图书馆自动化教程》的 ISBN 号 7-301-03210-2,  $S = 10 \times 7 + 9 \times 3 + 8 \times 0 + 7 \times 1 + 6 \times 0 +$

$5 \times 3 + 4 \times 2 + 3 \times 1 + 2 \times 0 = 130$ ,  $C = 11 - 130 \bmod 11 = 2$ , 与原校验码相符。

## 3 ISBN 校验的运行机制

一般情况下, ISBN 校验法是能够保证录入数据的准确性的。当然, 如果录入的时候不小心把上下两条书目的 ISBN 号调了一下, 校验的结果显示是正确的, 但实际上出错了, 即这种方法不能检查出在 ISBN 号之外出现的差错, 只能显示这些数字作为个体的内部联系。外部出错这种情况在本文中不予考虑。

模取 11 对 ISBN 校验意义重大。还是以 7-301-03210-2 为例, 第 8 位为 1, 权数为 3, 在别的数字输入正确的情况下, 这一位的输入有误, 假如为 4 或 5, 都会使  $S \bmod 11$  发生变化, 从而使校验位与实际得到结果有出入。可以得出的结果是, 只要第 8 位不为 1, 同时取值范围为在 0 到 9 中变化, 那么, 在其它 8 位数字不变的情况下, 都可以得到与校验位不同的结果, 即能够检查出错误。当然, 第 8 位的性质决定了它的取值只能在 0 到 9 中变化。就宏观而言, 还是以 7-301-03210-2 为例, 第三位为 0, 假如输入有误, 设为  $x$  ( $x$  不等于 0), 如果出现模相同的情况, 那么可以得到公式  $(130 + 8x) \bmod 11$ , 在 0 到 9 的范围内  $x = 0$  具有唯一性, 与题设不附, 所以假设不成立, 一定会有模不同的结果, 出错的时候也就一定会检查出来。模取 11 是在实际应用过程中经过比较后确立的, 假设模取 10, 那么,  $(130 + 8x) \bmod 10$ ,  $x = 5$  和 0 都会使最后的结果为 0, 结果不唯一, 导致

收稿日期: 2003-03-03

检验的时候出错。

以上的分析解释了 ISBN 校验的运行机理。录入人员在录入的时候,由于整天对着屏幕,可能会敲错数字,但是,ISBN 校验能够保证在输入数字中只错一个的时候检查出错来,从而加以更正,而如果错两个或多个数字,在一般情况下,ISBN 校验还是能够检查出来,只有在少数情况下,这种检验会出错,下文将会分析出错的原因。由于有了上述运行机制的保证,在长期的实践工作中,在很大概率上,ISBN 校验能够保证录入数据的准确性,因而深受广大图书馆工作者的好评。

#### 4 ISBN 校验的局限性

4.1 加和导致的错误。以 7-301-03210-2 为例,假设输入 7-301-03014,得到  $C = 11 - S \bmod 11 = 2$ ,实际上,两个 ISBN 是不同的,而计算机检不出来,这就涉及到一个由加和导致的错误的问题。

7-301-03210

10987 65432

上一行为 ISBN 号,下一行为权值,第 7 位和第 9 位对应的数字 2 和 0,权值为 4 和 2,那么,  $2 \times 4 + 0 \times 2 = 8$ ,假如输入有误,设输入的分别为  $x, y$ ,则  $4x + 2y = 8$ ,得到  $x = 0, y = 4; x = 2, y = 0$ ,第二个解符合原 ISBN 号,但是当输入的号中的第 7 位和第 9 位为第一个解中的两个数时,计算机是不能够检查出错的。当然,不只两位,三位或者更多位的加和也是有可能出现不唯一的情况,可能导致计算机检验结果出错。

4.2 权值导致的错误。权值为 10、9、8、7、6、5、4、3、

2,可以观察到  $5 + 6 = 4 + 7$ ,那么,在这些权值所对应的数字上同时加上或者减去一个整数,所得到的 C 会一样,计算机检验结果出错。例如,7-301-03210 与 7-302-01342310,所得到的 C 都为 2。当然,权值中  $2 + 4 + 5 = 3 + 8$ ,在这些权值对应的数字上同时加上或者减去一个整数,所得到的 C 也会一样,即权值导致的错误是无限定位数的,可以是两位相加,也可以是三位或者更多。实际上,权值导致的错误是加和导致的错误的情况中的一种,只是更有规律性。

4.3 另外还存在着一种可能,  $11 - S \bmod 11 = 11$ ,即  $S \bmod 11 = 0$ ,也就是得到的 S 为 11 的倍数,那么,在  $C = 11$  没有相应的规定码作为标识的时候,系统出错。当然,这种可能已经引起重视并有了相应的解决措施,可以通过改变 ISBN 算法进行补救,即将原公式稍微作一下改动得到  $C = 11 - [(S-1) \bmod 11 + 1]$  则可避免这种可能。

#### 5 余论

ISBN 校验的局限性源于该程序的简单性,通过对程序的修正,是可以修正这一不足的。当然,就目前的情况而言,ISBN 校验法基本上可以满足查找错误代码的要求,出错的概率也不大,是一种很好的方法。而到了一定阶段,随着文献数量的上升,出错的概率会增大。为了提高图书馆自动化系统的质量,笔者认为应该对这个问题加以重视。

#### 参考文献:

傅守灿等.图书馆自动化基础教程[M].北京:北京大学出版社,1996.87-100.

## The Work of ISBN Checkout and Its Localization

YU XUN-pei

(The Department of Information Management, Peking University, Beijing 100871, China)

**Abstract:** The article introduces the background of ISBN Checkout, and then explains the course and the function of it, indicates the mistake brought by summation or proportionment, brings forward that we should pay attention to it.

**Key words:** ISBN checkout; code; proportionment; procedure

#### 作者简介:

余训培,男,北京大学信息管理系本科生,北京大学中国经济研究中心学士,发表论文二十余篇。