ISBN Numbers

An Error-Detecting Code

Erik Fogwell and Nicole Fogwell
Advisor: Artem Zvavitch

Background

The International Standard Book Number (ISBN) is a unique numeric commercial book identifier. An ISBN is assigned to each edition and variation of a book. Books published before 2007 have a 10 digit ISBN, and books published before 2007 have a 13 digit ISBN. However, there is not a book for every 10 digit number. Only certain combinations of 10 or 13 digits are valid ISBN numbers. This is the beauty of the concept- when placing an order for a book by entering its ISBN, the computer can immediately tell whether the number is valid or not by using a division algorithm. The computer multiplies the first digit of the ISBN by 1, the second digit by 2, etc. It then sums each of these numbers together for a total. If the ISBN is 10 digits long, this total must be divisible by 11. If it is 13 digits long, the total must be divisible by 10. Thus, if the total proves to be not divisible by the set number, the computer knows that this ISBN is invalid and asks the user to re-enter it. This protects against both accidental swapping of digits and a single incorrect digit, both of which are common human errors.

ISBN 0-13-146413-2

An example ISBN code

ISBN DIGIT	MULTIPLIER	PRODUCT
0	1	0
1	2	2
3	3	9
1	4	4
4	5	20
6	6	36
4	7	28
1	8	8
3	9	27
2	10	20
	SUM:	154/11=14

Using the example 10-digit ISBN code, we can quickly use the ISBN test equation to see if the ISBN code is valid.

The original 10-digit ISBN code format, which was based on the work of Irish professor Gordon Foster, began to see widespread use as a way to label and identify books in 1970. There are 4 separate parts to each book's ISBN code, which are often separated by hyphens. The first section indicates the area that the publisher is located in. The second part represents the book's publishing company. The third number string is unique to each book that the company prints. For example, the first number (0) in the example code that was taken from my statistics book indicates that it was published in English. The 13 indicates that it was published by Prentice Hall and 146413 is the code for the book. While this is an effective way to organize and identify books, the unique part of the code is the last digit. Known as the check digit, it can be any number from 0-9 or X (which is treated as 10). The check number is calculated to make the following equation where x_n represents the nth digit of the ISBN code true:

$$X_1 + 2X_2 + 3X_3 + 4X_4 + 5X_5 + 6X_6 + 7X_7 + 8X_8 + 9X_9 + 10X_{10} = (11a)$$

Where a is a varying positive integer. Interestingly, the formula also works when the multipliers are reversed:

$$10X_1 + 9X_2 + 8X_3 + 7X_4 + 6X_5 + 5X_6 + 4X_7 + 3X_8 + 2X_9 + X_{10} = (11a)$$

\In words, when you multiply each digit of an ISBN number by its position in the code and add them together, the sum will be a multiple of 11. This means that when a book has been assigned the first 9 digits of its ISBN, the check digit for any 9-digit code is calculated by solving the equation for x_{10}. Just this alone is an effective way to organize the hundreds of thousands of books that are published each year. The code allows for up to 1 billion unique titles to be published and given a number to be organized by the system. But what makes the code even better is that though it is simple, you can use it to check if a code is invalid. The major benefit is for people ordering books using it's ISBN code. If the person placing the order makes a very simple mistake by mistyping a digit of the code, the error is able to be detected easily because the sum of all weighted digits is always divisible by 11 and if one digit is miscopied the code won't be valid. This can be proven by letting a represent the digit's weight and x and y be 2 arbitrary digits between 0 and 9.

$$ax = ay \ (\pm 11)$$
$$ax - ay = (\pm 11)$$
$$x - y = (\pm 11)/a$$

If x and y are not equal, there is no solution for this equation. The code also protects against accidentally swapping 2 adjacent digits and can be proven by letting x and y represent 2 adjacent digits and a being the multiplication weight of the first digit.

$$ax + (a+1)y = (a+1)x + ay \ (\pm 11)$$
$$ax + ay + y = ax + x + ay \ (\pm 11)$$
$$X = y \ (\pm 11)$$
$$x - y = (\pm 11)$$

These two simple checks alone have likely prevented countless book ordering mistakes and saved a lot money that would have been wasted on misprintings.

In 2007, the ISBN code received an update that made it much easier to check, at a cost. At the bottom of the example picture, you will see the updated version. The 9 major digits of the code remain the same, so there was no need to majorly overhaul the codes of books that were published before the change. The difference between the old and new codes is that all codes are now prefixed by the number 987 and the check digit is in most cases different. The new way to check if the code is valid is by plugging the digits into the following expression:

$$X_1 + 3X_2 + X_3 + 3X_4 + X_5 + 3X_6 + X_7 + 3X_8 + X_9 + 3X_{10} + X_{11} + 3X_{12} + X_{13}$$

In this case, however, the sum of this expression will always be divisible by 10 if it is a valid ISBN code. This adjusted method has a few advantages over the one that was used previously. Multiples of 10 are much easier to recognize than multiples of 11 and calculating the check digit was made easier. The check digit can be easily calculated by summing the first 12 digits. The check digit will be the number between 0 and nine that makes the total sum divisible by 12. Though it became easier to check, the new formula does not have a complete safeguard from people accidentally swapping 2 adjacent digits. This can be see by letting two digits be represented by a and b, respectively:

$$3a+b=3b+a (\pm 10)$$
$$2a-2b=(\pm 10)$$
$$a-b=(\pm 5)$$

This means that if the 2 digits have a difference of 5 (for example 6 and 1), the check will not always catch an accidental swap as the sum will still be some multiple of 10. However, the other safeguard still works:

$$3a=3b\pm 10 \quad a=b\pm 10$$
$$3a-3b=\pm 10 \quad a-b=\pm 10$$
$$a-b=\pm 10/3$$

Which both have no solution when a≠b and both a and b are integers between 0 and 9.

A test of the new ISBN format:

DIGIT	MULTIPLIER	PRODUCT
9	1	9
7	3	21,
8	1	8
0	3	0
1	1	1
3	3	9
1	1	1
4	3	12
6	1	6
4	3	12
1	1	1
3	3	9
1	1	1
	SUM:	90/10=9