

Codificación de información mediante códigos de barras*

L. HERNÁNDEZ¹ Y A. MARTÍN²

¹Departamento de Tratamiento de la Información y Codificación,
Instituto de Física Aplicada,
Consejo Superior de Investigaciones Científicas

²Departamento de Matemática Aplicada,
Universidad de Salamanca

`luis@iec.csic.es, delrey@usal.es`

Resumen

Se presentan en este artículo varios sistemas de codificación de información mediante códigos de barras. En particular se detalla la forma de elaborar el código de barras EAN13, posiblemente el más utilizado en la actualidad, y su relación con otro de los códigos más empleados: el ISBN para la catalogación de libros. Se incluyen y comentan los procesos matemáticos empleados para elaborar tales códigos.

Palabras clave: *Codificación de la información, códigos de barras, códigos detectores de errores, EAN13, ISBN.*

Clasificación por materias AMS: *94A15, 94B05, 94B60*

1 Introducción

El desarrollo tecnológico actual, la proliferación de los ordenadores y periféricos, así como la facilidad para el establecimiento de redes locales y su rapidez de acceso, ha permitido el desarrollo de nuevas aplicaciones matemáticas e informáticas a diferentes campos de la vida cotidiana. Una de ellas es la codificación de objetos de toda índole; en la que además de las herramientas informáticas que agilizan el tratamiento de la información, intervienen diferentes aspectos matemáticos, como son los algoritmos utilizados en la elaboración de los códigos y la verificación de que éstos son leídos correctamente.

* Parcialmente subvencionado por el Ministerio de Ciencia y Tecnología, TIC2001–0586.
Fecha de recepción: 23 de julio de 2003

Se denomina *código* a todo sistema de signos o señales y reglas que cambia la forma de un mensaje. Esta codificación se lleva a cabo de forma estándar y no secreta, es decir, su finalidad es la de resumir determinados datos y permitir una manipulación electrónica de los mismos. Desde este punto de vista, no es objetivo de un código impedir que la información que contiene sea accesible a personas diferentes de quienes elaboraron dicho código. No se trata pues de códigos secretos, que entran en el campo de la criptografía (para temas criptográficos véanse, por ejemplo, [10, 14]).

En particular, los códigos de barras que se presentan en este artículo son *no significativos* en el sentido de que no proporcionan más información que la contenida en el propio código, es decir, no permiten ocultar información, dado que su finalidad no es otra que la de transformar la información constituida por caracteres en una información gráfica, que puede ser tratada por medios informáticos. Por este motivo, no es conveniente que información confidencial se codifique por medio de códigos de barras.

Hoy en día, la mayor parte de los productos manufacturados, desde libros a latas de tomate, pasando por prendas de vestir, medicamentos, o paquetes de envío urgente, llevan una etiqueta con determinados símbolos o barras, que codifican información relativa a dicho artículo y que permiten identificarlo de forma unívoca. Así, se entiende por *código de barras* a un conjunto de líneas y números asociados a ellas, que va impreso en los productos de consumo y que se utiliza para su gestión informática. La información contenida en el código de barras hace referencia a datos relevantes del artículo, como el país de fabricación, su tamaño, propiedades, precio, etc. Estos datos son accesibles por medio de un lector óptico que “lee” el contenido del código mediante un rayo láser. La “lectura” es transformada por el software correspondiente y manipulada conforme a determinados requerimientos informáticos. El código leído es enviado a una base de datos que responde con el nombre del artículo, su precio y otros datos. En el caso de una venta, el artículo es dado de baja en el almacén, con lo que es posible gestionar las ventas diarias, el stock, etc.

En el presente trabajo se comentan diferentes métodos de codificación mediante códigos de barras. En §2 se presentan algunas definiciones básicas sobre códigos en su sentido matemático. En la sección 3 se hace un breve repaso a los códigos de barras más empleados. El caso particular del código de barras EAN13 se comenta con detalle en la sección 4. Finalmente, en §5 se presenta el código ISBN, su relación con el EAN13 y algunas de sus propiedades para detectar determinados tipos de errores.

2 Definiciones y Notación

Se llama *alfabeto* a un conjunto finito de q símbolos, $\Sigma_q = \{s_1, s_2, \dots, s_q\}$. Los elementos (vectores) de Σ_q^n se denominan *palabras de longitud n* , y se llama *código de longitud n* a todo subconjunto $\mathcal{C} \subset \Sigma_q^n$. Los elementos de \mathcal{C} se conocen como *codewords* o *palabras codificadas* (para un estudio más detallado véanse, por ejemplo, [5, 9, 15]). Si $q = 2$, los códigos se llaman *códigos binarios*.

La *distancia de Hamming* entre dos vectores $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$ de Σ_q^n se define como el número de posiciones en las que difieren dichos vectores y viene dada por:

$$d(x, y) = |\{i: 1 \leq i \leq n \text{ y } x_i \neq y_i\}|.$$

Dado un código \mathcal{C} , se define su *distancia mínima*, d , por la expresión

$$d = \min\{d(x, y): x, y \in \mathcal{C} \text{ y } x \neq y\}.$$

Si \mathcal{C} tiene k elementos, $|\mathcal{C}| = k$, el código \mathcal{C} suele llamarse un (n, k, d) -código. Los códigos con distancia mínima $d \geq 2$ pueden ser empleados para detectar $d - 1$ errores (*códigos detectores de errores*). Sea $d = 2e + 1 \geq 3$, entonces \mathcal{C} puede utilizarse para corregir errores (*código corrector de errores*), de modo que puede corregir $e = (d - 1)/2$ errores. Si $d = 2e \geq 4$, el código podrá corregir $e - 1 = (d - 2)/2$ errores. La capacidad de detectar y de corregir errores depende del valor de d y del algoritmo que se utilice. Así, un código con distancia $d = 5$, por ejemplo, puede utilizarse para detectar 4 errores, para corregir 2 errores, o para detectar 3 y corregir 1, etc.

Se define la *tasa de información* de un código \mathcal{C} como la razón entre la cantidad de información significativa de cada palabra y la longitud de cada palabra. Esta tasa puede determinarse mediante la expresión:

$$R = \log_q |\mathcal{C}|/n. \quad (1)$$

El concepto de tasa de información es natural en el sentido de que para codificar 4 palabras usando un alfabeto binario, bastaría con emplear las palabras 00, 01, 10 y 11; pero si se emplea un código de 4 palabras de longitud 3, entonces, por la expresión (1), su tasa de información sería $2/3$. Se llama *peso de Hamming* de un vector $x \in \Sigma_q^n$, y se representa por $w(x)$, al número de elementos no nulos que contiene, es decir, a la distancia de x al vector 0. Dado que la distancia entre dos vectores, x, y , es el número de posiciones en las que difieren, se tiene $d(x, y) = w(x - y)$. El *peso mínimo* de un código es el mínimo de los pesos de Hamming de las palabras no nulas de dicho código.

Los ejemplos más útiles de alfabetos son $\Sigma_q = \mathbb{Z}_q$ y, si q es la potencia de un número primo, el cuerpo $\Sigma_q = \mathbb{F}_q$ (recuérdese que \mathbb{Z}_q es un cuerpo si y sólo si q es primo). Un código \mathcal{C} es un *código lineal* sobre \mathbb{F}_q si es un subespacio vectorial de \mathbb{F}_q^n , y se dice que es un $[n, k]$ -código si dicho código tiene dimensión k . Si, además, su mínima distancia es d , entonces se llama un $[n, k, d]$ -código. Es claro que la tasa de información de un $[n, k, d]$ -código es $R = k/n$. Además como los códigos lineales son subespacios vectoriales, este peso mínimo coincide con la distancia mínima del código.

Para ilustrar las anteriores definiciones, en el Cuadro 1 se presentan dos ejemplos de códigos binarios.

\mathcal{C}_1 es un código binario de longitud 5 y se forma añadiendo un bit de paridad al final de cada una de las palabras de \mathbb{F}_2^4 ; es decir, al final de cada palabra se añade el bit 0 si la suma de sus bits es par, y el 1 en caso contrario. Claramente

Palabra	Código \mathcal{C}_1	Código \mathcal{C}_2
0000	00000	0000000
0001	00011	1010101
0010	00101	0110011
0011	00110	1100110
0100	01001	0001111
0101	01010	1011010
0110	01100	0111100
0111	01111	1101001
1000	10001	1111111
1001	10010	0101010
1010	10100	1001100
1011	10111	0011001
1100	11000	1110000
1101	11011	0100101
1110	11101	1000011
1111	11110	0010110

Cuadro 1: Ejemplos de dos códigos lineales binarios

este código puede detectar hasta un bit de error, pero no puede corregir errores dado que su distancia mínima es 2. \mathcal{C}_1 es un $[5, 4, 2]$ -código binario. Por su parte, el código \mathcal{C}_2 es de longitud 7 y distancia mínima 3. Dado que es un $[7, 4, 3]$ -código, es capaz de corregir hasta un bit de error.

Esta ganancia a la hora de corregir errores y no sólo de detectarlos tiene un precio. En este caso, según la expresión (1), la tasa de información de \mathcal{C}_2 es $4/7$; mientras que la tasa correspondiente a \mathcal{C}_1 es $4/5$. Como es de suponer, los códigos que atraen más atención son aquellos que poseen una alta tasa de información y proporcionan, además, una alta capacidad para corregir errores.

Los códigos de barras que se presentarán de forma detallada en este artículo son códigos detectores de errores puesto que en su elaboración se incluye un dígito de control que permite llevar a cabo tal detección; pero no permiten corregir errores y no son códigos lineales.

La combinación de dos códigos con distancia $d = 2$ permite obtener códigos con distancia mayor; por lo que pueden corregir errores. Es el caso del código binario del Servicio Postal americano utilizado para indicar el código postal en las cartas, llamado *código Postnet*. El hecho de que permita corregir errores se debe a que se puede determinar la posición en la que se ha producido el error. El código Postnet está formado por dos códigos de distancia $d = 2$: uno decimal de longitud 10 y otro binario de longitud 5 (que permite escribir cada dígito mediante barras). Este segundo código se muestra en el Cuadro 2, donde los ceros se representan por medio de barras cortas y los unos mediante barras largas. Ambos códigos se combinan para obtener un código binario de longitud 50 y distancia $d = 4$.

por un láser lineal, y basta con leer cualquier línea transversal al código, dado que todas ellas son iguales. No obstante, existen diferentes formas de llevar a cabo esta codificación. En general, la elección de un determinado tipo de código de barras depende de la aplicación para la que se desee utilizar.

Los códigos de barras suelen tener dos representaciones: una determinada cantidad de dígitos o caracteres en su parte inferior (no siempre presente) y una parte gráfica formada por unas barras verticales de diferente grosor y separadas por unos espacios paralelos. Los códigos codifican determinado *juego de caracteres*, es decir, un conjunto específico de letras, números y símbolos. Todo código de barras posee unos elementos característicos. Así, los *separadores de inicio y de fin* de cada código son combinaciones específicas de barras y espacios que indican al lector óptico dónde empieza y dónde termina el código a leer. Los códigos se llaman *bidireccionales* si pueden ser leídos tanto de derecha a izquierda como de izquierda a derecha. A continuación se presentan los códigos de barras lineales más utilizados en la actualidad, así como una breve descripción de cada uno de ellos (para más detalles ver [2, 3, 16]).

El *Código 39* (Código 3 de 9) es un código de longitud variable, adecuado para codificar datos alfanuméricos de carácter general. Cada carácter se representa mediante cinco barras y cuatro espacios, con sólo dos anchuras posibles: gruesa y fina (ver Figura 1(a)). El *Código 2 de 5* (Intercalado 2 de 5, ITF ó I-2/5) es un sistema de codificación (véase la Figura 1(b)) de propósito general para datos numéricos.

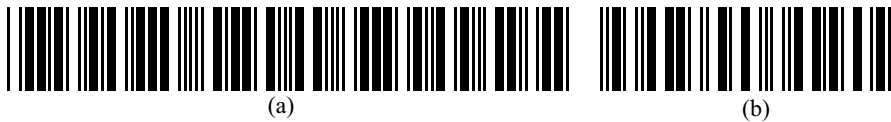


Figura 1: Ejemplos de Código 39 y Código 2 de 5

Codabar es un código para la codificación de números que incluye algunos caracteres especiales (ver Figura 2(a)). Dispone de 4 separadores de inicio/fin, que pueden llevar información adicional. El *Código 128*, más condensado que el Código 39, permite codificar 128 caracteres ASCII (se pueden utilizar diferentes juegos de caracteres para codificar caracteres extendidos que no pertenezcan al inglés). Cada carácter se representa mediante 3 barras y 3 espacios, que pueden tener cuatro anchuras diferentes (ver Figura 2(b)).



Figura 2: Ejemplos de Codabar y de Código 128

Existen otros códigos, como el *Código 93* que ofrece una densidad de información alfanumérica mayor que la que ofrecen los Códigos 39 y 128, incluyendo el código ASCII completo. Por su parte, el *Código de canal* es una familia de símbolos de barras unidimensionales diseñada para codificar cadenas de entre 2 y 7 dígitos con la menor longitud de símbolo posible.

Hay muchas aplicaciones en las que no se pueden emplear los códigos de barras presentados hasta ahora. Es el caso, por ejemplo, de la codificación que precise de los 256 caracteres ASCII de forma simultánea. Por este motivo se ha desarrollado la denominada *simbología 93i*, que codifica bytes y permite, por tanto, codificar todos los caracteres pertenecientes al código ASCII completo. La simbología 93i ([2]) extiende la del Código 39, incorpora los 65536 caracteres Unicode, su longitud es variable y utiliza 2 caracteres para comprobar los símbolos ó 6 códigos Reed-Solomon correctores de errores ([15]).

Dado que en ocasiones se requiere que el código de barras no supere determinada superficie, se ha desarrollado otra simbología que permite llevar a cabo este objetivo y se conoce como *Simbología para reducir espacio* o RSS (Reduced Space Symbology). Esta simbología consta de tres familias diferentes: *RSS-14*, *RSS Limitada* y *RSS Expandida*, que pueden ser utilizadas por el sistema EAN.UCC (European Article Numbering [7] y Uniform Code Council [17], respectivamente). Existe otra versión de esta simbología, conocida como *RSS-14 Apilada*, que corta un RSS-14 demasiado ancho y lo apila en dos filas.

4 El código de barras EAN13

Sin duda, el código de barras más extendido en el mundo es el llamado *EAN13* (existe una versión reducida y menos utilizada llamada *EAN8*). Este código data de 1977 y es para Europa el análogo al código UPC de Estados Unidos (Universal Product Code, [17]) o al código JAN de Japón (Japanese Article Number [13]). El EAN13 es una extensión del UPC, por lo que ha heredado algunas de sus características y arbitrariedades, como se mencionará más adelante. El nombre del código, EAN, procede de las iniciales de European Article Numbering ([7]), mientras que el 13 hace referencia al número de dígitos que van impresos en la parte inferior del código, de los que las barras son su representación gráfica. Como ya se mencionó, el EAN13 es *no significativo*, es decir, no proporciona ninguna información adicional que no esté contenida en los 13 dígitos que lo acompañan.

4.1 Determinación de los dígitos

Los 13 dígitos del código EAN13, c_1, \dots, c_{13} , se calculan como sigue:

1. *Código del país*: Son los dos primeros dígitos (algunos países utilizan 3), c_1 y c_2 , y son asignados por la organización nacional del sistema EAN a la que se ha adscrito la empresa que elabora el producto (ver [1]). Por ejemplo, los dos dígitos para España son 84, para el Reino Unido, 50, mientras que Alemania usa del 400 al 440.

2. *Código de la empresa*: Los siguientes 5 u 8 dígitos, c_3, \dots, c_{10} , se reservan para las diferentes empresas registradas en cada país.
3. *Código del producto*: Son los dígitos que restan hasta 12 (entre 2 y 5), c_8, \dots, c_{12} , y están a disposición del propietario de la marca.
4. *Dígito de control*: El último dígito, c_{13} , se calcula a partir de los 12 dígitos anteriores y permite decidir si el código de barras se ha leído correctamente. Su expresión simplificada es:

$$c_{13} = - \sum_{j=1}^6 (c_{2j-1} + 3c_{2j}) \pmod{10}. \quad (2)$$

Una vez que los 13 dígitos han sido determinados, cada uno de ellos es representado mediante una colección de barras y espacios verticales, y todo ello constituye el código EAN13.

4.2 Fundamentos matemáticos

Antes de proceder a la determinación de las barras y espacios del código, se justificarán las razones matemáticas de la elección del dígito de control dado en la expresión (2).

En general, el cálculo de un dígito de control, a_n , para una colección de $n-1$ dígitos dados: a_1, \dots, a_{n-1} , se lleva a cabo de manera que una operación sobre todos los dígitos de identificación: a_1, \dots, a_{n-1}, a_n , permita decidir si el código ha sido leído correctamente. Para ello se considera el conjunto de los enteros módulo k , \mathbb{Z}_k , y n aplicaciones $\sigma_i: \mathbb{Z}_k \rightarrow \mathbb{Z}_k$, y se supone que el código se ha leído correctamente si y sólo si se verifica la siguiente condición:

$$\sum_{i=1}^n \sigma_i(a_i) \equiv 0 \pmod{k}.$$

(Nótese que la elección del 0 en la congruencia anterior es arbitraria, es decir, no hay ninguna razón que impida elegir cualquier otro entero de \mathbb{Z}_k). La secuencia $\sigma_1, \dots, \sigma_n$ se denomina *esquema de control de dígitos*. Normalmente, las aplicaciones σ_i se definen por $\sigma_i(a_i) = \omega_i \cdot a_i$, con $\omega_i \in \mathbb{Z}_k$. Un *error de un único dígito* (ED para abreviar), es decir, confundir el dígito a_i con a'_i ; es detectable si y sólo si se verifica que $\sigma_i(a_i) \not\equiv \sigma_i(a'_i) \pmod{k}$. Por otra parte, un *error de transposición* (ET para abreviar), esto es, intercambiar el dígito a_i por el a_j ; es detectable si y sólo si $\sigma_i(a_i) + \sigma_j(a_j) \not\equiv \sigma_i(a_j) + \sigma_j(a_i) \pmod{k}$. En el caso en que se considere $k > 10$, se deben introducir caracteres extra que sustituyan a los valores $10, \dots, k-1$. El principal resultado sobre la detección de errores se basa en el siguiente

Teorema 1 *Si los números de identificación a_1, \dots, a_n , verifican*

$$(\omega_1, \omega_2, \dots, \omega_n) \cdot (a_1, a_2, \dots, a_n) = \sum_{i=1}^n \omega_i \cdot a_i \equiv 0 \pmod{k}, \quad (3)$$

entonces, un error de un único dígito es indetectable si y sólo si $\omega_i (a_i - a'_i) \equiv 0 \pmod k$; mientras que un error de transposición que intercambia los elementos de las posiciones i y j es indetectable si y sólo si $(\omega_i - \omega_j) (a_i - a_j) \equiv 0 \pmod k$.

La demostración de este Teorema es inmediata, basta tener en cuenta que cuando se comete un ED, la diferencia entre el valor correcto (esto es, cuando se utiliza a_i) para la expresión (3) y el valor incorrecto (cuando se emplea a'_i) es precisamente $\omega_i (a_i - a'_i)$. Por otra parte, la diferencia entre el valor correcto y el incorrecto para (3) cuando se lleva a cabo un ET es $(\omega_i - \omega_j) (a_i - a_j)$.

A partir del Teorema 1 si un dígito de control, a_n , verifica la condición (3) (en realidad esta condición puede ser contrastada si se consideran varios dígitos de control y no sólo uno), entonces, es inmediato determinar las condiciones de los pesos w_i de cada a_i , de modo que se pueda asegurar qué errores son detectables (ver Cuadro 3). Nótese que las unidades de \mathbb{Z}_k son los elementos $z \in \mathbb{Z}_k$ tales que $\text{mcd}(z, k) = 1$.

Error	Forma	Condición sobre el módulo k
ED	$a_i \rightarrow a'_i$	$\text{mcd}(w_i, k) = 1$
ET 1	$a_i \dots a_j \rightarrow a_j \dots a_i$	$\text{mcd}(w_j - w_i, k) = 1$
ET 2	$a_{i-1} a_i a_{i+1} \rightarrow a_{i+1} a_i a_{i-1}$	$\text{mcd}(w_{i+1} - w_{i-1}, k) = 1$
Gemelos 1	$\begin{matrix} a & a & \rightarrow & b & b \\ & i+1 & & i+1 \end{matrix}$	$\text{mcd}(w_i + w_{i+1}, k) = 1$
Fonética	$a_i 0 \leftrightarrow 1 a_i$	$j w_{i+1} \not\equiv (j-1) w_i \pmod k, 0 \leq j \leq k-1$
Gemelos 2	$\begin{matrix} a & c & a & \rightarrow & b & c & b \\ i-1 & i+1 & & & i-1 & i+1 \end{matrix}$	$\text{mcd}(w_{i-1} + w_{i+1}, k) = 1$

Cuadro 3: Condiciones del módulo k sobre diferentes errores

Si por ejemplo, $k = 10$ y los w_j son impares, y por tanto $w_i - w_j$ es par, las condiciones presentadas en la última columna del Cuadro 3 son incompatibles, es decir, no pueden verificarse todas simultáneamente. Así pues, no se pueden detectar, en el 100% de los casos, los errores mencionados anteriormente para el caso $k = 10$. De forma más general se verifica el siguiente

Teorema 2 *Si un esquema de detección de errores con un módulo par detecta los errores de un único dígito, entonces para cualesquiera i y j , existe un error de transposición que afecta a las posiciones i y j , que no puede ser detectado.*

También en este caso la demostración es inmediata puesto que si el esquema detecta todos los errores de tipo ED, entonces las aplicaciones σ_i son permutaciones de \mathbb{Z}_{2m} . Además, para detectar todos los errores ET que involucran a las posiciones i y j , es necesario que $\sigma_i(a) + \sigma_j(b) \neq \sigma_i(b) + \sigma_j(a)$, para todos los $a \neq b \in \mathbb{Z}_{2m}$. De donde se tiene que $\sigma(z) = \sigma_i(z) - \sigma_j(z)$ es una permutación de \mathbb{Z}_{2m} . Sumando los elementos de \mathbb{Z}_{2m} , módulo $2m$, se tiene que

$$m = m + 0 + 1 + (2m - 1) + 2 + (2m - 2) + \dots + (m - 1) + (m - 1),$$

y además

$$m = \sum z = \sum \sigma(z) = \sum \sigma_i(z) - \sum \sigma_j(z) = m - m = 0,$$

lo que es una contradicción.

De lo dicho anteriormente se puede afirmar que el dígito de control, c_{13} , dado por (2) para el código EAN13 está bien definido puesto que verifica

$$\sum_{j=0}^6 c_{2j+1} + 3 \sum_{j=1}^6 c_{2j} \equiv 0 \pmod{10}. \quad (4)$$

Nótese que para el EAN13, $w_j = 1$ si j es impar y $w_j = 3$ si j es par. El hecho de asignar diferentes pesos a las posiciones pares e impares hace que se puedan detectar (ver Teorema 1 y Cuadro 3) el 100% de los errores de un único dígito dado que $\text{mcd}(w_j, 10) = 1$, aunque no se puedan corregir. Además, no detecta la totalidad de los errores de transposición, tal y como se desprende del Teorema 2. En efecto, si se transponen los dígitos a_i y a_j , con i y j de diferente paridad, se tiene que la diferencia de las sumas dadas por (4) para cada uno de los dos casos es $(3a_i + a_j) - (3a_j + a_i) = 2(a_i - a_j)$, y el error es indetectable si $|a_i - a_j| = 5$.

Por ejemplo, si se tienen los 13 dígitos siguientes para un código EAN13: 8412345678905, es inmediato comprobar que tal código es correcto. En efecto, basta con verificar que se cumple la congruencia dada en (4):

$$(8 + 1 + 3 + 5 + 7 + 9 + 5) + 3(4 + 2 + 4 + 6 + 8 + 0) \pmod{10} = 110 \pmod{10} \equiv 0.$$

Ahora bien, si en lugar de considerar el código anterior, se ha cometido un ED al considerar el código 8412395678905, se puede detectar que éste es erróneo puesto que se tiene

$$(8 + 1 + 3 + 5 + 7 + 9 + 5) + 3(4 + 2 + 9 + 6 + 8 + 0) \pmod{10} \equiv 5 \pmod{10},$$

pero no hay forma de saber dónde se cometió el error, y éste no puede corregirse. El código original podría haber sido 3412395678905 ó 8412895678905, etc.

Por otra parte, si se ha producido un ET entre las posiciones sexta y décimo primera y los 13 dígitos considerados fueran 8412395678405, éste código se daría por correcto, a pesar de haberse producido un error, dado que también se verifica que

$$(8 + 1 + 3 + 5 + 7 + 4 + 5) + 3(4 + 2 + 9 + 6 + 8 + 0) \pmod{10} \equiv 0 \pmod{10}.$$

Para determinar el porcentaje de errores de transposición de dos dígitos consecutivos que son detectados por el EAN13, basta tener en cuenta que de los 90 posibles ET de esta forma, todos se pueden detectar salvo que se tengan los siguientes pares de dígitos consecutivos: 05, 16, 27, 38, 49 y sus inversos, es decir, se puede detectar el $80/90 = 89,89\%$ de estos errores. Razonando de forma análoga se obtiene el mismo valor para todos los ET.

Sin embargo, si los pesos dados a las posiciones fueran $w_j = 1$ si j es impar y $w_j = 2$ si j es par, éste esquema detectaría el 100% de los ET, dado que en este caso si se transponen, por ejemplo, los dígitos consecutivos a y b , se tiene que el error sólo es indetectable si $(2a + b) - (2b + a) = a - b \equiv 0 \pmod{10}$ (ver [11] para esquemas con tres pesos diferentes). Sin embargo, en este caso no se detectarían todos los ED.

4.3 Determinación de las barras y espacios

Para determinar el código de barras bajo el que se escribirán los 13 dígitos de que consta, se utiliza el *módulo* como unidad básica de representación y corresponde a la mínima anchura que puede tener una barra ($\langle 1 \rangle$) o un espacio ($\langle 0 \rangle$). De esta forma la codificación $\langle 000 \rangle$ hace referencia a un espacio de módulo 3, mientras que la codificación $\langle 11 \rangle$ se refiere a una barra de módulo 2. Además, cada código tiene siempre tres separadores (ver Figura 3), que son los mismos para todos los códigos EAN13, y que indican dónde comienza el código (*separador izquierdo* o de inicio), dónde termina (*separador derecho* o de fin) y cuál es su centro (*separador central*). Los separadores derecho e izquierdo son siempre $\langle 101 \rangle$ mientras que el separador central es $\langle 01010 \rangle$. Estos valores permiten al escáner que lee las barras determinar la anchura utilizada por cada módulo del código, lo que posibilita que un mismo escáner sea capaz de leer códigos de barras de diferentes tamaños. Las barras de dichos separadores son ligeramente más largas que las restantes del código.

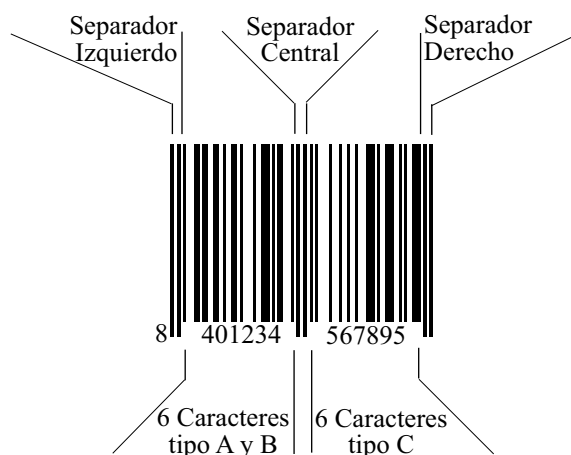


Figura 3: Descripción de las partes de un código EAN13

El primer dígito de los 13 del EAN13 se coloca en la parte exterior izquierda del código, es decir, antes del separador izquierdo y no se representa gráficamente, aunque influye en la representación gráfica de los dígitos que ocupan las posiciones #2 a #7, como se verá posteriormente. Los restantes 12 dígitos se separan en dos grupos de 6 dígitos cada uno, de modo que el primer grupo se coloca y codifica entre el separador izquierdo y el central; mientras que el segundo grupo queda situado entre el separador central y el derecho.

La representación gráfica de cada dígito está formada por dos barras y dos espacios y ocupa siempre 7 módulos, es decir, cada carácter está codificado por un vector binario de longitud 7 y peso de Hamming variable —el número de 1's que contiene—. Así, el vector $\langle 0111001 \rangle$ correspondería a un espacio de módulo

1, una barra de módulo 3, un espacio de módulo 2 y una barra de módulo 1. La codificación anterior también podría haberse llevado a cabo mediante los cuatro dígitos que indican los módulos de cada espacio y barra: 1321.

El número total de módulos de un código EAN13 es, por tanto, de $2 \cdot 3 + 12 \cdot 7 + 5 = 95$, si bien se acostumbra a dejar zonas mudas alrededor del mismo para facilitar su lectura por el escáner. Por otra parte, como la anchura estándar de cada módulo es de $0,33 \text{ mm.}$ y su altura estándar de $22,85 \text{ mm.}$, resulta que el tamaño estándar para un código de este tipo, incluidas las zonas mudas, es de $37,29 \times 26,26 \text{ mm.}$ (se permite un escalado entre 0,8 y 2 veces el valor estándar anterior).

Las representaciones gráficas de los 6 primeros dígitos siempre empiezan por un espacio (es decir, por uno o varios 0's) y terminan por una barra (esto es, por uno o varios 1's); mientras que para el segundo grupo es al revés: empiezan por una barra y terminan por un espacio. Nótese que esta distribución es compatible con la codificación de los separadores izquierdo, derecho y central. El primer grupo utiliza dos patrones diferentes (tipos *A* y *B*) para codificar cada dígito, mientras que el segundo grupo utiliza un tercer patrón (tipo *C*). La codificación del primer grupo (posiciones #2 a #7) se hace utilizando el Cuadro 4, según el valor del primer dígito del código. Por esta razón no es necesario representar el primer dígito del código mediante barras, sin que se pierda por ello la información que el mismo proporciona.

Primer dígito	#2	#3	#4	#5	#6	#7
0	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>
1	<i>A</i>	<i>A</i>	<i>B</i>	<i>A</i>	<i>B</i>	<i>B</i>
2	<i>A</i>	<i>A</i>	<i>B</i>	<i>B</i>	<i>A</i>	<i>B</i>
3	<i>A</i>	<i>A</i>	<i>B</i>	<i>B</i>	<i>B</i>	<i>A</i>
4	<i>A</i>	<i>B</i>	<i>A</i>	<i>A</i>	<i>B</i>	<i>B</i>
5	<i>A</i>	<i>B</i>	<i>B</i>	<i>A</i>	<i>A</i>	<i>B</i>
6	<i>A</i>	<i>B</i>	<i>B</i>	<i>B</i>	<i>A</i>	<i>A</i>
7	<i>A</i>	<i>B</i>	<i>A</i>	<i>B</i>	<i>A</i>	<i>B</i>
8	<i>A</i>	<i>B</i>	<i>A</i>	<i>B</i>	<i>B</i>	<i>A</i>
9	<i>A</i>	<i>B</i>	<i>B</i>	<i>A</i>	<i>B</i>	<i>A</i>

Cuadro 4: Patrones de codificación para las posiciones #2 a #7

Así por ejemplo, si el primer dígito es 3, los dígitos de las posiciones #2, #3 y #7 se codifican según el patrón *A*, mientras que los de las posiciones #4, #5 y #6 utilizan el patrón *B*. Finalmente, es necesario conocer cómo se codifica cada dígito para cada uno de los patrones. Esta codificación se presenta en el Cuadro 5, de modo que las dos barras (*b*) y los dos espacios (*e*) de cada patrón están caracterizados como sigue:

- Patrón *A*. Su secuencia es *ebeb* y los vectores binarios de este tipo son de longitud 7 y peso impar, esto es, 3 ó 5.

- Patrón *B*. Su secuencia es *ebeb* y los vectores binarios pertenecientes a este tipo son de longitud 7 y peso par, es decir, 2 ó 4.
- Patrón *C*. Su secuencia es *bebe* y sus vectores binarios son de longitud 7 y peso par: 2 ó 4 (al ser su paridad diferente de la de *A*, que es el patrón con el que siempre empieza el código, éste puede leerse bidireccionalmente).

Dígito	Patrón <i>A</i>	Patrón <i>B</i>	Patrón <i>C</i>
0	3211 → ⟨0001101⟩	1123 → ⟨0100111⟩	3211 → ⟨1110010⟩
1	2221 → ⟨0011001⟩	1222 → ⟨0110011⟩	2221 → ⟨1100110⟩
2	2122 → ⟨0010011⟩	2212 → ⟨0011011⟩	2122 → ⟨1101100⟩
3	1411 → ⟨0111101⟩	1141 → ⟨0100001⟩	1411 → ⟨1000010⟩
4	1132 → ⟨0100011⟩	2311 → ⟨0011101⟩	1132 → ⟨1011100⟩
5	1231 → ⟨0110001⟩	1321 → ⟨0111001⟩	1231 → ⟨1001110⟩
6	1114 → ⟨0101111⟩	4111 → ⟨0000101⟩	1114 → ⟨1010000⟩
7	1312 → ⟨0111011⟩	2131 → ⟨0010001⟩	1312 → ⟨1000100⟩
8	1213 → ⟨0110111⟩	3121 → ⟨0001001⟩	1212 → ⟨1001000⟩
9	3112 → ⟨0001011⟩	2113 → ⟨0010111⟩	3112 → ⟨1110100⟩

Cuadro 5: Asignación de los patrones de codificación

Es claro que si se elige otra codificación diferente a la presentada en los Cuadros 4 y 5, es decir, otra reordenación en los patrones para cada dígito inicial, o las asignaciones a dígitos dentro de cada patrón de codificación, se obtendrán diferentes códigos de barras. En este caso no se estaría utilizando el estándar EAN13, sino otra codificación diferente.

4.4 EAN13 y códigos binarios

Después de lo mencionado acerca de la representación mediante espacios y barras (alternados) los dígitos 0–9 en un código EAN13, parece necesario dar una justificación matemática a dicho proceso, aunque el inventor del código UPC, precursor del EAN13, no se basara en conceptos matemáticos.

En primer lugar, es claro que la representación de cada dígito necesita, al menos, de 2 espacios y 2 barras, puesto que como el final de la representación de un dígito debe servir para marcar el inicio de la representación del dígito siguiente, no pueden emplearse 2 espacios y 1 barra (o viceversa). Por tanto, el código binario a emplear debe ser de longitud mayor o igual que 4.

Por otra parte, se ha de garantizar, también, que cada dígito pueda ser codificado de dos maneras diferentes: una para el patrón *A* (vectores de peso impar) y otra para el *B* (vectores de peso par), por lo que hace falta un mínimo de 20 codewords. Además, todas las representaciones de los dígitos deben empezar por ⟨0⟩ (espacio) y terminar por ⟨1⟩ (barra). Dado que sólo existen $2^4 = 16$, codewords de longitud 6 de la forma

$$\langle 0 \ x_1 \ x_2 \ x_3 \ x_4 \ 1 \rangle, \quad x_i \in \mathbb{F}_2,$$

es necesario utilizar codewords de longitud, al menos, 7. Supongamos entonces que los codewords son de longitud 7 y que de los $2^7 = 128$ posibles codewords sólo se consideran los $2^5 = 32$ de la forma preestablecida, es decir,

$$\langle 0 \ x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ 1 \rangle, \quad x_i \in \mathbb{F}_2.$$

De estos 32 codewords se deben eliminar todos aquellos para los que no haya exactamente 3 cambios de la forma 01 ó 10, dado que los 2 espacios y las 2 barras de que consta cada representación (*ebeb*) deben estar alternados. Una vez eliminados los que no verifican esta condición, quedan los siguientes posibles codewords:

$$\begin{aligned} &0000101, \ 0001001, \ 0001011, \ 0001101, \ 0010001, \\ &0010011, \ 0010111, \ 0011001, \ 0011011, \ 0011101, \\ &0100001, \ 0100011, \ 0100111, \ 0101111, \ 0110001, \\ &0110011, \ 0110111, \ 0111001, \ 0111011, \ 0111101. \end{aligned}$$

Como se puede observar, la mitad de los codewords anteriores son de peso impar (3 ó 5) y la otra mitad son de peso par (2 ó 4), por lo que dicha división puede utilizarse como criterio para separar los dos patrones y obtener dos códigos binarios de longitud 7, diferentes:

$$\begin{aligned} \mathcal{C}_A &= \{0001101, 0011001, 0010011, 0111101, 0100011, \\ &\quad 0110001, 0101111, 0111011, 0110111, 0001011\}, \\ \mathcal{C}_B &= \{0100111, 0110011, 0011011, 0100001, 0011101, \\ &\quad 0111001, 0000101, 0010001, 0001001, 0010111\}. \end{aligned}$$

La ordenación de los codewords (que corresponde a la presentada en el Cuadro 5) es la misma que la empleada en el código UPC americano, inventado en 1973 por George J. Laurer —4 años antes que el EAN13—, quien, en comunicación personal, nos ha asegurado que la ordenación del patrón *A* (y por tanto la asignación de los dígitos 0–9 para el primer grupo de 6 caracteres) la llevó a cabo de forma completamente arbitraria (lo mismo que la ordenación de los patrones *A* y *B* del Cuadro 4). Por otra parte, la ordenación del patrón *B* la tomó, de alguna manera, como imagen especular del patrón *A*. A partir de las asignaciones mencionadas, la correspondiente al patrón *C* es inmediata de modo que permita decidir al escáner si la lectura del código se lleva a cabo de derecha a izquierda o al revés.

$$\begin{aligned} \mathcal{C}_C &= \{1110010, 1100110, 1101100, 1000010, 1011100, \\ &\quad 1001110, 1010000, 1000100, 1001000, 1110100\}. \end{aligned}$$

Así pues, los códigos \mathcal{C}_A , \mathcal{C}_B y \mathcal{C}_C son códigos binarios de longitud 7 y distancia mínima 2, es decir son (7, 10, 2)-códigos, pero no son códigos lineales, dado que no son subespacios vectoriales de \mathbb{F}_2^7 . Su tasa de información es:

$$R = \frac{\log_2 10}{7} = 0,475.$$

El hecho de añadir barras al código EAN13 para representar dígitos, dota a este de cierta capacidad para corregir errores, de manera análoga a como lo hace el código Postnet, ya comentado, aunque esta posibilidad no se utiliza habitualmente. Por ejemplo, si se pierde un dígito del código o es leído erróneamente —ya sea porque hay manchas o roturas o porque la paridad no es la adecuada—, el error es detectado porque el EAN13 puede detectar un error al ser su distancia mínima $d = 2$. Además, se puede saber la posición del error a partir de la lectura de las barras. Conocida ésta, el dígito puede ser recuperado sin más que utilizar la expresión (4). Por otra parte, si el error se comete en la lectura del primer dígito, que no está representado mediante barras, también puede ser corregido debido a la paridad de los 6 dígitos siguientes: basta con observar la asignación de los patrones del Cuadro 4.

4.5 Ejemplo de código

Para determinar el código EAN13 de un producto ficticio se calculan los 13 dígitos de que consta y se determina el código de barras correspondiente.

Si la empresa utiliza el sistema EAN de la organización AECOC (Asociación Española de Codificación Comercial [1]) de España, los dos primeros dígitos serán 84, es decir, $c_1 = 8$, $c_2 = 4$. Si la AECOC ha asignado a esta empresa los dígitos 765432, éstos serán los siguientes, es decir, $c_3 = 7$, $c_4 = 6$, $c_5 = 5$, $c_6 = 4$, $c_7 = 3$ y $c_8 = 2$. Si 0209 son los 4 dígitos que el propietario de la marca asigna al artículo ficticio, entonces $c_9 = 0$, $c_{10} = 2$, $c_{11} = 0$ y $c_{12} = 9$. Según esto, los dígitos para el código de barras son: 847654320209 c_{13} . Para calcular el dígito de control se procede como se indica en la expresión (2):

$$c_{13} = -(23 + 3 \cdot 27) \pmod{10} = -104 \pmod{10} = 6.$$

Por lo que los dígitos para el código de barras son: 8476543202096.

A continuación, se determinará la forma en que los dígitos calculados se representan mediante barras y espacios verticales para dar lugar al código de barras EAN13.

Como el primer dígito del código es 8, la secuencia de los patrones A y B es A, B, A, B, B, A (ver Cuadro 4). Dado que los 6 primeros dígitos son 476543, su codificación, según el Cuadro 5, es la siguiente:

$$\begin{aligned} 4 &\rightarrow \langle 0100011 \rangle, & 7 &\rightarrow \langle 0010001 \rangle, & 6 &\rightarrow \langle 0101111 \rangle, \\ 5 &\rightarrow \langle 0111001 \rangle, & 4 &\rightarrow \langle 0011101 \rangle, & 3 &\rightarrow \langle 0111101 \rangle. \end{aligned} \quad (5)$$

Procediendo de forma análoga para el segundo grupo de 6 dígitos, para los que se utiliza el patrón C , se obtiene su codificación:

$$2 \rightarrow \langle 1101100 \rangle, \quad 0 \rightarrow \langle 1110010 \rangle, \quad 9 \rightarrow \langle 1110100 \rangle, \quad 6 \rightarrow \langle 1010000 \rangle. \quad (6)$$

Finalmente, la representación del código de barras EAN13 para el artículo que se está considerando es la que se muestra en la Figura 4. Dicha figura se ha elaborado colocando en primer lugar el separador izquierdo $\langle 101 \rangle$, a continuación

los vectores binarios del primer grupo presentado en (5), luego el separador central $\langle 01010 \rangle$, a continuación los vectores binarios del segundo grupo mostrado en (6), y finalmente, el separador derecho $\langle 101 \rangle$.

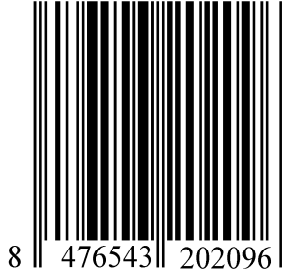


Figura 4: Ejemplo de Código EAN13

Con relación a la posibilidad de corregir errores del EAN13, si en el código anterior se hubiera perdido el primer dígito, observando que los pesos de los 6 dígitos siguientes son; 3, 2, 5, 4, 4, 5, se deduce que el orden de los patrones es: A, B, A, B, B, A , por lo que a partir del Cuadro 4 se tiene que dicho dígito es 8 y el error se habría corregido. Por otra parte, si, por ejemplo, el dígito que ocupa la sexta posición (el 4), no pudiera leerse o fuera leído como $\langle 0111101 \rangle$, se tendría que, o bien no se conoce su valor o se sabría que es erróneo puesto que su peso es impar, cuando debería ser par. En cualquiera de los dos casos su valor real se podría conocer sin más que utilizar la expresión (4). En efecto, si se denota por x al valor desconocido o erróneo, se tendría que

$$(8 + 7 + 5 + 3 + 0 + 0 + 6) + 3(4 + 6 + x + 2 + 2 + 9) \equiv 0 \pmod{10},$$

es decir,

$$\begin{aligned} 3x &\equiv -29 - 69 \pmod{10} \equiv -98 \pmod{10} = 2, \\ x &\equiv 2 \cdot 3^{-1} \pmod{10} \equiv 2 \cdot 7 \pmod{10} = 4. \end{aligned}$$

4.6 EAN13 y las particiones de 7

Para concluir esta sección se presentará otra explicación a la señalada anteriormente para la elaboración de las barras y espacios que forman parte de un código EAN13, que tiene como base las particiones de un entero.

Dado que cada dígito del EAN13 ocupa un módulo de 7 unidades —suma de las anchuras de 2 barras y 2 espacios—, el problema consiste en escribir el número 7 como suma de 4 dígitos de diferentes formas. La descomposición de un número entero como suma de dígitos se conoce como *particiones de un entero*, y es un tema ampliamente estudiado en Combinatoria (ver, por ejemplo, [4, Capítulo 5.3] y [6, Capítulo 2.7]). En particular, las particiones de 7 como suma

de dígitos son las siguientes:

$$\begin{array}{ll}
 7 & = 1 + 1 + 1 + 1 + 1 + 1 + 1 & = 1 + 1 + 1 + 1 + 1 + 2 \\
 & = 1 + 1 + 1 + 2 + 2 & = 1 + 1 + 1 + 1 + 3 \\
 & = 1 + 2 + 2 + 2 & = 1 + 1 + 2 + 3 \\
 & = 1 + 1 + 1 + 4 & = 1 + 3 + 3 \\
 & = 1 + 2 + 4 & = 1 + 1 + 5 \\
 & = 2 + 2 + 3 & = 3 + 4 \\
 & = 2 + 5 & = 1 + 6.
 \end{array}$$

De todas ellas, sólo son interesantes para el EAN13 las que contienen exactamente 4 dígitos: (1, 2, 2, 2), (1, 1, 2, 3) y (1, 1, 1, 4). Por otra parte, dado que el orden de colocación de las barras y espacios hace que el código sea visualmente diferente, se deben considerar todas las formas de escribir los tres grupos de cuatro dígitos anteriores, es decir, sus permutaciones:

$$\begin{array}{llllll}
 [1, 2, 2, 2], & [2, 1, 2, 2], & [2, 2, 1, 2], & [2, 2, 2, 1], & & \\
 [1, 1, 2, 3], & [1, 1, 3, 2], & [1, 2, 1, 3], & [1, 2, 3, 1], & [1, 3, 1, 2], & [1, 3, 2, 1], \\
 [2, 1, 1, 3], & [2, 1, 3, 1], & [2, 3, 1, 1], & [3, 1, 1, 2], & [3, 1, 2, 1], & [3, 2, 1, 1], \\
 [1, 1, 1, 4], & [1, 1, 4, 1], & [1, 4, 1, 1], & [4, 1, 1, 1]. & &
 \end{array}$$

Se observa que la mitad de las 20 permutaciones obtenidas son simétricas respecto de las restantes. Por ejemplo, si se consideran los 10 grupos:

$$\begin{array}{llllll}
 [2, 1, 2, 2], & [2, 2, 2, 1], & [1, 1, 3, 2], & [1, 2, 1, 3], & [1, 2, 3, 1], \\
 [1, 3, 1, 2], & [3, 1, 1, 2], & [3, 2, 1, 1], & [1, 1, 1, 4], & [1, 4, 1, 1], &
 \end{array} \quad (7)$$

los restantes 10 son su simétricos. Por tanto, se puede considerar como patrón *A* al determinado por la codificación de las 10 permutaciones dadas en (7), y como patrón *B* a las restantes 10 permutaciones. Además, como el patrón *A* comienza por barra y termina por espacio, se puede utilizar el mismo conjunto de permutaciones para el patrón *C*, dado que éste empieza por espacio y concluye con barra. Así pues, asignando una permutación de las dadas en (7) a cada uno de los 10 dígitos, se llega a la codificación de cada patrón dada en el Cuadro 5.

Nótese que si las 10 permutaciones que definen el patrón *A* o si su orden fuera diferente del mostrado en el Cuadro 5, se obtendrían códigos de barras diferentes para una misma cadena de dígitos y, consecuentemente, no pertenecerían al estándar EAN13.

5 Los códigos ISBN y EAN13

Como es sabido, el código ISBN (International Standar Book Number, [12]) es un sistema internacional de numeración e identificación de libros surgido en 1965, y aunque no sea un código de barras propiamente dicho, se incluye en este artículo porque es posible elaborar un código EAN13 para cada código ISBN, de modo que pueda ser tratado informáticamente.

El código ISBN contiene 10 dígitos divididos en cuatro grupos por tres guiones. Dado que las posiciones de los guiones son irrelevantes para el código, se hará caso omiso de las mismas. El último dígito es el de control (la letra X si su valor es 10) y se calcula mediante la siguiente expresión:

$$b_{10} = - \sum_{i=1}^9 p_i \cdot b_i \pmod{11}, \quad (8)$$

donde $p_i = 11 - i$ es el peso de cada posición. La verificación de la lectura del código se lleva a cabo de forma análoga a como se hizo con el código EAN13, es decir, se comprueba si se verifica la siguiente congruencia:

$$\sum_{i=1}^{10} p_i \cdot b_i = 10b_1 + 9b_2 + \dots + 2b_9 + b_{10} \equiv 0 \pmod{11}. \quad (9)$$

Es fácil probar que el código ISBN detecta el 100% de los ED's. En efecto, si al cambiar el dígito b_i por b'_i el error no fuera detectable, los valores de las dos sumas de comprobación dadas por (9) serían múltiplos de 11, por lo que también lo sería su diferencia: $(11-i)(b_i - b'_i)$, lo cual es imposible puesto que 11 es primo y $|b_i - b'_i| < 11$. Este código también detecta el 100% de los ET's. En efecto, razonando como antes, la diferencia de las sumas dadas por (9) sería múltiplo de 11, pero esta diferencia es $(p_i b_i + p_j b_j) - (p_j b_i + p_i b_j) = (j - i)(b_i - b_j)$, que tampoco puede ser múltiplo de 11. De forma similar se puede probar que es capaz de detectar el 100% de los errores Gemelos 2 dados en el Cuadro 3, aunque no detecta todos los errores Gemelos 1 ni Fonéticos. Así pues, desde el punto de vista de la detección de errores, puede afirmarse que el código ISBN es preferible al EAN13 (véase [8] para un estudio sobre la detección de errores dobles en ISBN). Para esta mejora es básico que se utilice el número 11, que es primo, y, por tanto, \mathbb{Z}_{11} es un cuerpo.

Para determinar el código de barras EAN13 a partir de un código ISBN dado $b_1 b_2 b_3 b_4 b_5 b_6 b_7 b_8 b_9 b_{10}$, basta con calcular los 13 dígitos que le corresponden en formato EAN13. En primer lugar se eliminan los guiones y el dígito de control: $b_1 b_2 b_3 b_4 b_5 b_6 b_7 b_8 b_9$. A continuación se añaden los dígitos 978 al inicio del código ISBN de modo que $c_1 = 9$, $c_2 = 7$, $c_3 = 8$, $c_i = b_{i-3}$, para $4 \leq i \leq 12$. Finalmente se determina el dígito de control como se señaló en (2):

$$c_{13} = - \sum_{j=1}^6 (c_{2j-1} + 3c_{2j}) \pmod{10} = - \left(8 + \sum_{j=2}^4 b_{2j} + 3 \sum_{j=1}^5 b_{2j-1} \right) \pmod{10}.$$

Si, por ejemplo, el código ISBN fuera 84-7897-421-0, bastará con escribir 978847897421 y calcular su dígito de control según (2) o (5):

$$c_{13} = -(8 + 21 + 3 \cdot 29) \pmod{10} = -116 \pmod{10} = 4.$$

Por lo que su código de barras EAN13 sería: 9788478974214.

Recíprocamente, para convertir un código EAN13, $987c_4, \dots, c_{13}$, en formato ISBN, se eliminan los tres primeros dígitos, 978, y el dígito de control, c_{13} ; y a continuación se determina el dígito de control ISBN según la expresión (8).

Como ejemplo, se puede tomar como código EAN13 el calculado anteriormente: 9788478974214, y proceder como se ha indicado: se eliminan los tres primeros dígitos y el último: 847897421, y se calcula el dígito de comprobación según la expresión (8):

$$-(80 + 36 + 56 + 56 + 54 + 35 + 16 + 6 + 2) \pmod{11} = -341 \pmod{11} = 0.$$

Con lo que el código ISBN es: 8478974210.

Hay otro código basado en la aritmética módulo 11, al igual que el ISBN, que sí puede detectar el 100 % de los errores señalados en el Cuadro 3; aunque no otros errores, como errores dobles de transposición o de dos dígitos. Es el código utilizado por Bancos y Cajas de Ahorros en los números de las cuentas bancarias. El código asigna pesos geométricos con base 2 a las posiciones de los dígitos, en lugar de pesos aritméticos:

$$(2, 2^2, 2^3, \dots, 2^{10}) \pmod{11} = (2, 4, 8, 5, 10, 9, 7, 3, 6, 1).$$

Las cuentas bancarias se identifican por 20 dígitos: los 4 primeros corresponden a la entidad bancaria, b_1, b_2, b_3, b_4 ; los 4 siguientes a la sucursal, s_1, s_2, s_3, s_4 ; el noveno dígito, c_1 , es el de control de los 8 primeros; mientras que el décimo, c_2 , es el de control de los 10 últimos, d_1, d_2, \dots, d_{10} , que constituyen el número de la cuenta. Los valores de c_1 y c_2 se determinan como sigue:

$$c_1 \equiv -(4b_1 + 8b_2 + 5b_3 + 10b_4 + 9s_1 + 7s_2 + 3s_3 + 6s_4) \pmod{11},$$

$$c_2 \equiv -(d_1 + 2d_2 + 4d_3 + 8d_4 + 5d_5 + 10d_6 + 9d_7 + 7d_8 + 3d_9 + 6d_{10}) \pmod{11}.$$

Sin embargo, si el dígito de control es 10, en lugar de cambiarlo por la letra X, se sustituye por el número 1, lo que supone una ambigüedad y, por tanto, disminuye la capacidad de este código para detectar errores.

Existen otros esquemas basados en la aritmética modular que, sorprendentemente, son bastante utilizados a pesar de que ni siquiera son capaces de detectar el 100 % de los errores de un único dígito. Son sistemas que utilizan los 10 dígitos en su código, pero hacen módulo 7 ó 9 para determinar el dígito de control. Es el caso de los cheques del Servicio Postal de Estados Unidos, los cheques de viaje de American Express, los servicios de mensajería de Federal Express y United Parcel Service, o los billetes de las compañías aéreas. Por ejemplo, el dígito de control de los cheques de viaje de American Express se determina de modo que sumado a los anteriores sea divisible por 9. Esto significa que un ED que involucre a los números 0 y 9 no será detectado. Por su parte, cuando se hace módulo 7 la efectividad para detectar ED's es ligeramente menor dado que no se detectan errores en los que intervengan los pares 0, 7; 1, 8 y 0, 9; aunque aumente la efectividad para detectar los ET's, que en ningún caso llega al 100 %.

Agradecimientos. Los autores agradecen al evaluador anónimo sus sugerencias para la mejora en la versión final de este artículo.

Referencias

- [1] Asociación Española de Codificación Comercial (AECOC). <http://www.aecoc.es/>
- [2] Association for Automatic Identification and Data Capture Technologies (AIM Global Network). <http://www.aimglobal.org/sitemap/>
- [3] American National Standard Intitute (ANSI). <http://www.ansi.org/>
- [4] C. Chuan Chong and K. Khee Meng. *Principles and techniques in combinatorics*. World Scientific, Singapore, 1992.
- [5] G.C. Clark, Jr. and J.B. Cain. *Error-correction coding for digital communications*. Plenum Press, New York, 1981.
- [6] G.M. Constantine. *Combinatorial theory and statistical design*. John Willey & Sons, New York, 1987.
- [7] European Article Numbering (EAN). <http://www.ean-int.org/index800.html>
- [8] L. Egghe and R. Rousseau. *The detection of double errors in ISBN- and ISSN-like codes*. Math. Comput. Modelling 33:943–955, 2001.
- [9] D. Hankerson, D.R. Hoffman, D.A. Leonard, C.C. Lindner, K.T. Phelps, C.A. Rodger and J.R. Wall. *Coding theory and cryptography, The essentials*. Marcel Dekker, Pure and Applied Mathematics, 234, 2nd edition, 2000.
- [10] A. Fúster Sabater, D. de la Guía Martínez, L. Hernández Encinas, F. Montoya Vitini y J. Muñoz Masqué. *Técnicas criptográficas de protección de datos*. RA-MA, Madrid, 2ª edición, 2000.
- [11] J.A. Gallian. *Error detecting methods*. ACM Computing Surveys 28(3):504–517, September 1996.
- [12] The International ISBN Agency. <http://www.isbn-international.org/>
- [13] Japanese Article Number (JAN). <http://www.n-barcode.com/shurui-en/jan.html>
- [14] A. Menezes, P. van Oorschot, and S. Vanstone. *Handbook of applied cryptography*. CRC Press, Boca Raton, FL, 1997.
- [15] W.W. Peterson and E.J. Weldon, Jr. *Error-correcting codes*. The MIT Press, Cambridge, 2nd edition, 1972.
- [16] L. Steen (Ed.). *Las Matemáticas en la vida cotidiana*. Addison-Wesley Iberoamericana España, S.A., Madrid, 3ª edición, 1999.
- [17] Uniform Code Council (UCC). <http://www.uc-council.org/>