

Cómo funcionan los identificadores inteligentes

Contenido

[Introducción](#)

[Números de tarjeta de crédito](#)

[Números de seguridad social de EE. UU.](#)

[Números CUSIP](#)

[Números de Ruteo ABA](#)

Introducción

Este documento describe identificadores inteligentes, que son patrones de escaneo de contenido integrados que detectan ciertos tipos de datos. Para esta versión, el sistema implementará identificadores inteligentes para números de tarjetas de crédito, números de la Seguridad Social de EE. UU., números CUSIP y números de routing ABA.

Internamente, un identificador inteligente consiste en una expresión regular que coincide con las cadenas de candidatos, junto con una función de validación que verifica la coincidencia de candidatos de alguna manera. Por ejemplo, la función de validación de un número de tarjeta de crédito garantiza que el dígito de verificación sea correcto.

Las expresiones regulares para cada identificador inteligente incluirán delimitadores de límite de palabra ('b') en ambos extremos. (Esto evita que el sistema coincida con un número de la seguridad social de EE. UU., por ejemplo, en medio de una cadena más larga de dígitos.) Por simplicidad, se omiten en las descripciones siguientes.

La implementación de identificadores inteligentes debe tener cuidado con las coincidencias superpuestas, porque es posible que una subcadena encontrada por la expresión regular no se valide. Por ejemplo, un filtro está buscando números de tarjeta de crédito con la cadena 999 4321 9999 9999 99995 1234 5678 9000 debe encontrar el número de tarjeta de crédito válido 4321 999 999 995, a pesar de que una simple búsqueda de expresiones regulares para los posibles números encontraría 9999 4321 999 99999 y 9995 1234 5678 9000.

Números de tarjeta de crédito

Un número de tarjeta de crédito comienza con un tipo de tarjeta de longitud variable, que indica si el número es VISA, MasterCard, AMEX, etc., y termina con un dígito de verificación. Los diferentes tipos de tarjeta utilizan números diferentes de dígitos en todo el número, pero el cálculo del dígito de verificación es el mismo en cada caso.

Tenga en cuenta que las tarjetas enRoute o JCB no coinciden. Además, los números VISA de 13 dígitos no existen y no coincidirán en nuestra implementación.

Los números de tarjeta de crédito de 16 dígitos coincidirán con una de las siguientes expresiones regulares:

[0-9]{4}-[0-9]{4}-[0-9]{4}-[0-9]{4}
[0-9]{4}\.[0-9]{4}\.[0-9]{4}\.[0-9]{4}
[0-9]{4} [0-9]{4} [0-9]{4} [0-9]{4}
[0-9]{16}

Con el prefijo siendo "4", "51"- "55" o "6011".

Los números AMEX de 15 dígitos coincidirán con una de las siguientes expresiones regulares:

[0-9]{4}-[0-9]{6}-[0-9]{5}
[0-9]{4}\.[0-9]{6}\.[0-9]{5}
[0-9]{4} [0-9]{6} [0-9]{5}
[0-9]{15}

Con los prefijos permitidos siendo "34" o "37".

Los números de 14 dígitos del Club Diners coincidirán con una de las siguientes expresiones regulares:

[0-9]{4}-[0-9]{6}-[0-9]{4}
[0-9]{4}\.[0-9]{6}\.[0-9]{4}
[0-9]{4} [0-9]{6} [0-9]{4}
[0-9]{14}

Con los prefijos permitidos siendo "300"- "305", "36", o "38".

Tenga en cuenta que las expresiones regulares definen un agrupamiento específico de dígitos para una longitud de tarjeta de crédito dada y que si hay puntuación entre los dígitos, debe ser el mismo en todo.

El último dígito de un número de tarjeta de crédito es un dígito de verificación creado mediante el algoritmo Luhn. Trabajando desde el extremo derecho del número, doble cada segundo dígito. A continuación, agregue los dígitos individuales de los números resultantes (tanto los que se duplicaron como los que no lo fueron). Si el resultado es un múltiplo de 10, el número es válido.

Por ejemplo, dado el número 1234 5678 9012 3456:

1 2 3 4 5 6 8 0 1 3 5 6

Doble: 2 2 6 4 10 6 14 8 18 0 2 6 4 10 6

Añadiendo 2 + 2 + 6 + 4 + 1 + 0 ... + 1 + 0 + 6 da 64, que no es un múltiplo de 10, por lo que el número no es válido.

Dado el número 1234 5678 9876 3333:

1 2 3 4 5 6 8 8 7 3 3 3 3

Doble: 2 2 6 4 10 6 14 8 18 8 14 6 3 3

Añadiendo $2 + 2 + 6 + 4 + 1 + 0 \dots + 6 + 3$ proporciona 80, que es un múltiplo de 10, por lo que el número es válido.

Números de seguridad social de EE. UU.

Los números de la seguridad social se dividen en un número de área de 3 dígitos, asignado geográficamente, un número de grupo de 2 dígitos asignado en un orden concreto dentro de un área y un número de serie de 4 dígitos asignado secuencialmente.

Nuestra implementación utilizará las siguientes expresiones regulares:

```
[0-9]{3}-[0-9]{2}-[0-9]{4}
[0-9]{3}\.[0-9]{2}\.[0-9]{4}
[0-9]{3} [0-9]{2} [0-9]{4}
```

A continuación se muestran algunos ejemplos de las expresiones anteriores:

```
555-55-5555
555.55.5555
555 55 5555
```

La Administración de Seguridad Social mantiene una lista de los números de área/grupo que se han asignado: SSN emitido [3]. Pero como este documento cambia periódicamente, no podemos confiar en él para su validación. La función de validación comprobará que ninguno de los 3 campos son ceros y que los primeros 3 dígitos son inferiores a 800. (La referencia anterior utiliza 771 como límite, pero el SSA ya ha asignado números con los primeros 3 dígitos 771 y 772.)

(Los números que comienzan por 666 no están asignados y los números del rango 987-65-4320 a 987-65-4329 están reservados para publicidad. Además, 078-05-1120 es el SSN más utilizado; era el SSN real de un secretario de una empresa de cartera, que usaba el número como ejemplo.)

Números CUSIP

Los números del CUSIP (Comité de Procedimientos Uniformes de Identificación de Seguridad) son 9 identificadores alfanuméricos que identifican valores norteamericanos de diversos tipos. El número se divide en un número de emisor de 6 caracteres, que identifica de forma única al emisor (por ejemplo, una empresa), un sufijo de 2 caracteres que identifica el valor concreto; Por ejemplo, acciones comunes frente a acciones preferidas frente a instrumentos de renta fija frente a opciones.

El código CUSIP Smart Identifier utilizará las siguientes expresiones regulares:

```
[0-9]{3}[0-9a-zA-Z]{3} [0-9a-zA-Z]{2} [0-9]
[0-9]{3}[0-9a-zA-Z]{3}-[0-9a-zA-Z]{2}-[0-9]
[0-9]{3}[0-9a-zA-Z]{3}[0-9a-zA-Z]{2}[0-9]
```

La función de validación es similar a la utilizada para los números de tarjetas de crédito. La única diferencia es que las letras del número CUSIP se convierten a un valor numérico asignando A=10, B=11, ..., Z=35.

Un ejemplo del sitio web cusip.com utiliza el número CUSIP 392690 QT 3:

3 9 2 6 9 0 Q 3

Convertir letras: 3 9 2 6 9 0 26 29 3

Doble: 3 18 2 12 9 0 26 58 3

Añadiendo $3 + 1 + 8 + 2 + 1 + 2 + \dots + 5 + 8 + 3$ da 50, que es un múltiplo de 10, por lo que el número original era válido.

Números de Ruteo ABA

Un número de routing ABA (American Banking Association) es un valor de 9 dígitos. Los primeros 4 dígitos son el símbolo de ruteo de la Reserva Federal, los 4 siguientes el identificador de la institución y los últimos un dígito de verificación.

El código de identificador inteligente del número de routing ABA utilizará las expresiones regulares siguientes:

```
[0-9]{4} [0-9]{4} [0-9]
[0-9]{4}-[0-9]{4}-[0-9]
[0-9]{9}
```

La función de validación implicaba multiplicar cada dígito por 3, 7, 1, ...; si la suma de los productos es un múltiplo de 10, el número es válido.

Por ejemplo, considere el número 123 456 789:

1 2 3 4 6 8 9

Multiplicar por: 3 7 1 3 1 3 7 1

Producto: 3 14 3 12 35 6 21 56 9

Añadir $3 + 14 + 3 + 12 + 35 + 6 + 21 + 56 + 9$ da 159, que no es un múltiplo de 10, por lo que el número original no era válido.

Dado el número 322 271 627:

3 2 2 7 1 6 7

Multiplicar por: 3 7 1 3 1 3 7 1

Producto: 9 14 2 6 49 1 18 14 7

La adición de $9 + 14 + 2 + 6 + 49 + 1 + 18 + 14 + 7$ da 120, que es un múltiplo de 10, por lo que el número original era válido.

(Aunque algunos rangos de símbolos de ruteo de la Reserva Federal están reservados, y por lo tanto no están asignados, el algoritmo de validación no comprobará los números reservados, para evitar tener que revisarlo si la ABA cambia su política.)