

# Funktionsweise von Smart Identifiers

## Inhalt

[Einführung](#)

[Kreditkartennummern](#)

[Sozialversicherungsnummern in den USA](#)

[CUSIP-Nummern](#)

[ABA-Weiterleitungsnummern](#)

## Einführung

In diesem Dokument werden intelligente Identifikatoren beschrieben, bei denen es sich um integrierte Content-Scanning-Muster handelt, die bestimmte Datentypen erkennen. Für diese Version implementiert das System Smart Identifier für Kreditkartennummern, US-Sozialversicherungsnummern, CUSIP-Nummern und ABA-Routing-Nummern.

Intern besteht ein Smart Identifier aus einem regulären Ausdruck, der Kandidaten-Zeichenfolgen zuordnet, sowie einer Validierungsfunktion, die die Übereinstimmung mit einem Kandidaten auf eine bestimmte Weise überprüft. Beispielsweise stellt die Validierungsfunktion für eine Kreditkartennummer sicher, dass die Prüfziffer korrekt ist.

Die regulären Ausdrücke für jede Smart Identifier enthalten auf beiden Seiten Wortbegrenzungsanker ('b'). (Dies verhindert, dass das System eine US-Sozialversicherungsnummer zuordnet, z. B. mitten in einer größeren Ziffernfolge.) Aus Gründen der Einfachheit werden diese in den nachfolgenden Beschreibungen weggelassen.

Bei der Implementierung von Smart Identifiers müssen Überlappungen berücksichtigt werden, da eine durch den regulären Ausdruck gefundene Unterzeichenfolge möglicherweise nicht validiert wird. Ein Filter sucht beispielsweise nach Kreditkartennummern unter der Zeichenfolge 999 4321 999 999 999 99995 1234 5678 900. Die gültige Kreditkartennummer 4321 9999 999999 sollte finden 995, obwohl eine einfache Prüfung der regulären Ausdrücke nach möglichen Nummern 999 4321 999 9999 und 9995 1234 5678 9000 finden würde.

## Kreditkartennummern

Eine Kreditkartennummer beginnt mit einem Kartentyp variabler Länge, der angibt, ob es sich um eine VISA, MasterCard, AMEX usw. handelt, und endet mit einer Prüfziffer. Verschiedene Kartentypen verwenden unterschiedliche Zahlen in der gesamten Zahl, aber die Berechnung der Prüfziffer ist in jedem Fall gleich.

Beachten Sie, dass die enRoute- oder JCB-Karten nicht übereinstimmen. Außerdem existieren 13-stellige VISA-Nummern nicht und werden bei der Implementierung nicht zugeordnet.

16-stellige Kreditkartennummern stimmen mit einem der folgenden regulären Ausdrücke überein:

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

Das Präfix lautet "4", "51"- "55" oder "6011".

Die 15-stelligen AMEX-Nummern stimmen mit einem der folgenden regulären Ausdrücke überein:

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

Die zulässigen Präfixe sind "34" oder "37".

Die 14-stelligen Diners Club-Nummern stimmen mit einem der folgenden regulären Ausdrücke überein:

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

Die zulässigen Präfixe sind "300"- "305", "36" oder "38".

Bitte beachten Sie, dass reguläre Ausdrücke eine bestimmte Gruppierung von Ziffern für eine bestimmte Kreditkartenlänge definieren und dass bei einer Zeichensetzung zwischen den Ziffern durchgängig die gleiche sein muss.

Die letzte Ziffer einer Kreditkartennummer ist eine Prüfziffer, die mit dem Luhn-Algorithmus erstellt wurde. Am rechten Ende der Nummer doppelte jede zweite Ziffer. Addieren Sie dann die einzelnen Ziffern der resultierenden Zahlen (sowohl die doppelten als auch die nicht). Wenn das Ergebnis ein Vielfaches von 10 ist, ist die Zahl gültig.

Beispiel: Nr. 1234 5678 9012 3456:

1 2 4 5 6 7 8 9 0 1 2 4 5 6

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

Hinzufügen  $2 + 2 + 6 + 4 + 1 + 0 \dots + 1 + 0 + 6$  gibt 64, was nicht ein Vielfaches von 10, so ist die Zahl ungültig.

Nr. 1234 5678 9876 3333:

1 2 4 5 6 7 8 9 7 6 3 3 3 3

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

Hinzufügen  $2 + 2 + 6 + 4 + 1 + 0 \dots + 6 + 3$  gibt 80, was ein Vielfaches von 10, so ist die Zahl

gültig.

## Sozialversicherungsnummern in den USA

Die Sozialversicherungsnummern werden in eine dreistellige Bereichsnummer unterteilt, die geografisch zugewiesen wird, eine zweistellige Gruppennummer, die in einer bestimmten Reihenfolge innerhalb eines Bereichs zugewiesen wird, und eine vierstellige Seriennummer, die sequenziell zugewiesen wird.

Bei unserer Implementierung werden folgende reguläre Ausdrücke verwendet:

```
[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}
```

Hier einige Beispiele für die obigen Ausdrücke:

```
555-55-5555
555,55,5555
555 555 555
```

Die Sozialversicherung führt eine Liste der Bereich-/Gruppennummern, die zugewiesen wurden: SSN ausgestellt [3]. Da dieses Dokument jedoch regelmäßig geändert wird, können wir uns nicht auf die Validierung verlassen. Die Validierungsfunktion überprüft, ob keines der drei Felder alle Nullen ist und ob die ersten drei Ziffern kleiner als 800 sind. (Bei der vorherigen Referenz wird 771 als Grenzwert verwendet, aber die SSA hat bereits Zahlen mit den ersten drei Ziffern 771 und 772 zugewiesen.)

(Nummern, die mit 666 beginnen, werden nicht zugewiesen, und Nummern im Bereich 987-65-4320 bis 987-65-4329 sind für Werbung reserviert. Außerdem ist 078-05-1120 die am häufigsten verwendete SSN. Es war die tatsächliche SSN eines Sekretärs in einem Geldbörsenunternehmen, die die Nummer als Beispiel verwendete.)

## CUSIP-Nummern

CUSIP (Committee on Uniform Security Identification Procedures)-Nummern sind 9 alphanumerische Kennungen, die nordamerikanische Wertpapiere verschiedener Art identifizieren. Die Zahl ist in eine 6-stellige Emittentenummer unterteilt, die den Emittenten (z. B. ein Unternehmen) eindeutig bezeichnet, eine 2-stellige Nachsilbe zur Kennzeichnung des jeweiligen Wertpapiers; z. B. Stammaktien im Vergleich zu Vorzugsaktien im Vergleich zu Optionen im Vergleich zu fest verzinslichen Instrumenten.

Der CUSIP Smart Identifier-Code verwendet folgende reguläre Ausdrücke:

```
[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]
```

Die Validierungsfunktion ähnelt der für Kreditkartennummern verwendeten Funktion. Der einzige

Unterschied besteht darin, dass Buchstaben in der CUSIP-Nummer durch Zuweisen von A=10, B=11, ..., Z=35 in einen numerischen Wert umgewandelt werden.

Ein Beispiel von der Website cusip.com verwendet die CUSIP-Nummer 392690 QT 3:

3 9 2 6 9 0 Q T 3

Buchstaben umwandeln: 3 9 2 6 9 0 26 29 3

Double: 3 18 2 12 9 0 26 58 3

Hinzufügen  $3 + 1 + 8 + 2 + 1 + 2 + 2 + \dots + 5 + 8 + 3$  gibt 50, was ein Vielfaches von 10, so war die ursprüngliche Nummer gültig.

## ABA-Weiterleitungsnummern

Eine ABA-Routing-Nummer (American Banking Association) ist ein 9-stelliger Wert. Die ersten vier Ziffern sind das Routing-Symbol der Federal Reserve, die nächsten vier die Instituts-ID und die letzte eine Prüfziffer.

Der Smart Identifier-Code der ABA-Routing-Nummer verwendet folgende reguläre Ausdrücke:

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

Die Validierungsfunktion beinhaltet das Multiplizieren jeder Ziffer mit 3, 7, 1, ...; Wenn die Summe der Produkte ein Vielfaches von 10 ist, ist die Zahl gültig.

Beispiel: 123 456 789:

1 2 4 5 6 7 9

Multiplizieren mit: 3 7 1 3 7 1 3 7 1

Produkt: 3 14 3 12 35 6 21 56 9

Das Hinzufügen von  $3 + 14 + 3 + 12 + 35 + 6 + 21 + 56 + 9$  ergibt 159, was nicht ein Vielfaches von 10 ist, sodass die ursprüngliche Zahl war ungültig.

Unter der Nummer 322 271 627:

3 2 2 7 1 6 2

Multiplizieren mit: 3 7 1 3 7 1 3 7 1

Produkt: 9 14 2 6 49 1 18 14 7

Das Hinzufügen von  $9 + 14 + 2 + 6 + 49 + 1 + 18 + 14 + 7$  ergibt 120, was ein Vielfaches von 10 ist, sodass die ursprüngliche Nummer war gültig.

(Obwohl einige Bereiche der Routingsymbole der Federal Reserve reserviert und daher nicht zugewiesen sind, überprüft der Validierungsalgorithmus nicht auf reservierte Nummern, um zu vermeiden, dass diese geändert werden müssen, wenn die ABA ihre Richtlinien ändert.)