

智慧識別符號的工作原理

目錄

[簡介](#)

[信用卡號](#)

[美國社會保險號碼](#)

[CUSIP編號](#)

[ABA路由編號](#)

簡介

本文檔介紹智慧識別符號，它是檢測特定型別資料的內建內容掃描模式。對於此版本，系統將實施信用卡號、美國社會保險號、CUSIP號和ABA路由號的智慧識別符號。

在內部，智慧識別符號由匹配候選字串的正規表示式以及以某種方式檢查候選匹配的驗證函陣列成。例如，信用卡號的驗證功能可確保校驗位正確。

每個智慧識別符號的正規表示式將包含位於兩端的單詞邊界錨點('b')。(這可以防止系統匹配美國社會保險號碼，例如，在較長的數字串中間。)為簡單起見，以下說明中省略了這些內容。

智慧識別符號實現必須小心重疊匹配，因為正規表示式找到的子字串可能無法驗證。例如，篩選器正在根據字串9999 4321 9999 999 999 9995 1234 5678 9000掃描信用卡號，應該找到有效的信用卡號4321 9999 999 999 9999995，即使對可能的數字進行簡單的正規表示式掃描會找到9999 4321 99 999和995 12 4 5678 9000。

信用卡號

信用卡號以可變長度卡型別開頭，表示該號碼是否是VISA、MasterCard、AMEX等，並以校驗數字結尾。不同的卡型別在整個數字中使用不同的位數，但是每種情況下檢查位的計算是相同的。

請注意，enRoute或JCB卡不匹配。此外，13位的VISA號碼不存在，並且在我們的實施中不會匹配。

16位數的信用卡號將匹配以下正規表示式之一：

```
[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}
```

首碼為「4」、「51」—「55」或「6011」。

15位的AMEX數字將與以下正規表示式之一匹配：

[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}

允許的字首為「34」或「37」。

14位的Diners Club編號將匹配以下正規表示式之一：

[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}

允許的字首為「300」 — 「305」、「36」或「38」。

請注意，正規表示式為給定的信用卡長度定義特定的數字分組，並且，如果數字之間有標點，則它必須始終相同。

信用卡號中的最後一個數字是使用Luhn演算法建立的校驗數字。從數字的右端開始，每第二個數字加倍。然後加上結果數字的個位數字（雙位數字和不雙位數字）。如果結果為10的倍數，則數字有效。

例如，給定數字1234 5678 9012 3456:

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6

雙 : 2 2 6 4 10 6 14 8 18 0 2 6 4 10 6

加上2 + 2 + 6 + 4 + 1 + 0 ... + 1 + 0 + 6得出64，這不是10的倍數，因此數字無效。

致電1234 5678 9876 3333:

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

雙 : 2 2 6 4 10 6 14 8 18 8 14 6 6 3 6

加上2 + 2 + 6 + 4 + 1 + 0 ... + 6 + 3得出80，這是10的倍數，因此數字是有效的。

美國社會保險號碼

將社會保障號碼劃分為地理上指定的3位數地區號、區域內按特定順序指定的2位數群組號、順序指定的4位數序列號。

我們的實現將使用以下正規表示式：

[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}

以下是上述表達式的一些示例：

555-55-5555
555.55.5555
555 55 5555

社會保險管理局維護已分配的區域/組編號清單：SSN發佈[3]。但是由於本文檔會定期更改，因此我們不能依靠它進行驗證。驗證功能將檢查3個欄位是否全部都是零，前3位是否小於800。（上一個參考使用771作為限制，但SSA已分配了前3位數字771和772的號碼。）

(以666開頭的號碼未分配，987-65-4320到987-65-4329範圍內的號碼保留用於廣告。此外，078-05-1120是最被誤用的SSN;這是一個錢包公司秘書的實際SSN，並以此數字為例。)

CUSIP編號

CUSIP(Committee on Uniform Security Identification Procedures)編號是9個字母數字識別符號，用於標識各種型別的北美證券。該數字被分為6個字元的發行者編號，其唯一地標識發行者（例如公司），2個字元的字尾標識特定的安全性；例如，普通股與優先股與優先股與期權與固定收入工具。

CUSIP智慧識別符號代碼將使用以下正規表示式：

$[0-9]^3[0-9a-zA-Z]^3 [0-9a-zA-Z]^2 [0-9]$
 $[0-9]^3[0-9a-zA-Z]^3-[0-9a-zA-Z]^2-[0-9]$
 $[0-9]^3[0-9a-zA-Z]^3[0-9a-zA-Z]^2[0-9]$

驗證功能類似於用於信用卡號的驗證功能。唯一的區別是，通過分配A=10, B=11, ..., Z=35將CUSIP編號中的字母轉換為數值。

cusip.com網站中的示例使用CUSIP編號392690QT 3:

3 9 2 6 9 0 Q T 3

轉換字母：3 9 2 6 9 0 26

雙：3 18 2 12 9 0 26

加上3 + 1 + 8 + 2 + 1 + 2 + ... + 5 + 8 + 3得出50，這是10的倍數，因此原始數字有效。

ABA路由編號

ABA（美國銀行協會）的路徑號碼是9位數的值。前4位是美聯儲路徑符號，後4位是機構識別符號，最後一位是校驗位。

ABA路由編號智慧識別符號代碼將使用以下正規表示式：

$[0-9]^4 [0-9]^4 [0-9]$
 $[0-9]^4-[0-9]^4-[0-9]$
 $[0-9]^9$

驗證函式涉及將每個數字乘以3、7、1、...;如果產品總數是10的倍數，則數字有效。

例如，考慮數字123 456 789:

1 2 3 4 5 6 7 8 9

乘以：3 7 1 3 7 1 3 7

產品:3 14 3 12 35 6 21

加上 $3 + 14 + 3 + 12 + 35 + 6 + 21 + 56 + 9$ ，得出159，這不是10的倍數，因此原始數字無效。

號碼是322 271 627:

3 2 2 7 1 6 2

乘以：3 7 1 3 7 1 3 7

產品:9 14 2 6 49 1 18 14

加上 $9 + 14 + 2 + 6 + 49 + 1 + 18 + 14 + 7$ ，得出120，這是10的倍數，因此原始數字有效。

(雖然某些範圍的聯邦儲備局路由符號是保留的，因此沒有分配，但驗證演算法不會檢查保留的號碼，以避免在ABA更改其策略時不得不修改它。)