

智能标识符的工作方式

目录

[简介](#)

[信用卡号](#)

[美国社会安全数字](#)

[CUSIP号码](#)

[ABA路由号](#)

简介

本文档介绍智能标识符，这些标识符是检测特定类型数据的内置内容扫描模式。对于此版本，系统将实施信用卡号、美国社会保障号、CUSIP号和ABA路由号的智能标识符。

在内部，智能标识符由匹配候选字符串的正则表达式以及以某种方式检查候选匹配的验证函数组成。例如，信用卡号的验证功能可确保校验位正确。

每个智能标识符的正则表达式在两端都包括字边界锚点('b')。（这可以防止系统匹配美国社保号码，例如，在较长的数字字符串中间。）为简单起见，以下说明中省略了这些内容。

智能标识符实现必须小心重叠匹配，因为正则表达式找到的子字符串可能无法验证。例如，过滤器正在根据字符串9999 4321 9999 9999 9995 1234 5678 9000扫描信用卡号，应找到有效的信用卡号4321 9999 99999995，即使简单的正则表达式扫描可能的数字，也会找到9999 4321 9999 9999和9995 1234 5678 9000。

信用卡号

信用卡号以可变长度卡类型开头，指示该号码是否是VISA、万事达卡、美国运通卡等，以支票数字结尾。不同的卡类型在整个数字中使用不同的位数，但校验位计算在每种情况下都相同。

请注意，enRoute或JCB卡不匹配。此外，13位VISA号码不存在，在我们的实施中将不匹配。

16位信用卡号将匹配以下正则表达式之一：

```
[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]{4} [0-9]{4}
[0-9]{16}
```

前缀为“4”、“51” — “55”或“6011”。

15位AMEX号码将匹配以下正则表达式之一：

[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}

允许的前缀为“34”或“37”。

14位数的Diners Club编号将匹配以下正则表达式之一：

[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}

允许的前缀为“300” — “305”、“36”或“38”。

请注意，正则表达式为给定信用卡长度定义了特定的数字分组，如果数字之间有标点符号，则在整个过程中必须相同。

信用卡号中的最终数字是使用Luhn算法创建的支票数字。从数字的右端工作，每秒两位。然后将产生的号码的各个数字相加（两个数字加倍，另一个数字不加倍）。如果结果是10的倍数，则数字有效。

例如，给定数字1234 5678 9012 3456:

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

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

添加2 + 2 + 6 + 4 + 1 + 0 ... + 1 + 0 + 6表示64，它不是10的倍数，因此该数字无效。

根据数字1234 5678 9876 3333:

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

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

添加2 + 2 + 6 + 4 + 1 + 0 ... + 6 + 3可得到80，这是10的倍数，因此该数字有效。

美国社会安全数字

社会保障号码分为3位区号，按地理位置分配，按区域内特定顺序分配的2位组号，按顺序分配的4位序列号。

我们的实现将使用以下正则表达式：

[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}

以下是上述表达式的一些示例：

555-55-5555
555.55.5555
555 55 5555

社会保障局维护已分配的区域/组号列表：SSN发布[3]。但是，由于本文档会定期更改，因此我们无法依靠它进行验证。验证功能将检查3个字段中是否没有全部为零，以及前3个数字小于800。（上一个引用使用771作为限制，但SSA已分配了前3个数字771和772的数字。）

(从666开始的号码未分配，范围从987-65-4320到987-65-4329的号码保留用于广告。此外，078-05-1120是最误用的SSN;这是某钱包公司秘书的实际SSN，以数字为例。)

CUSIP号码

CUSIP (统一安全识别程序委员会) 编号是9个字母数字标识符，用于识别各种类型的北美证券。该数字被划分为一个6个字符的颁发者编号，该编号唯一地标识颁发者（例如公司），一个2个字符的后缀，用于标识特定的安全性；例如，普通股、优先股与期权、固定收益工具。

CUSIP智能标识符代码将使用以下正则表达式：

```
[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]
```

验证功能类似于信用卡号。唯一的区别是，CUSIP编号中的字母通过指定A=10、B=11、...、Z=35来转换为数值。

cusip.com网站的一个示例使用CUSIP编号392690 QT 3:

3 9 2 6 9 0 Q T 3

转换字母：3 9 2 6 9 0 26 29 3

双倍：3 18 2 12 9 0 26 58 3

添加3 + 1 + 8 + 2 + 1 + 2 + ... + 5 + 8 + 3可得到50，即10的倍数，因此原始数有效。

ABA路由号

ABA (美国银行协会) 路由号是9位数值。前4位是美联储路由符号，后4位是机构标识符，最后一位是支票数字。

ABA路由号智能标识符代码将使用以下正则表达式：

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

验证函数涉及将每个数字乘以3、7、1、...;如果产品总和是10的倍数，则该数字有效。

例如，请考虑数字123 456 789:

1 2 3 4 5 6 7 8 9

乘以 : 3 7 1 3 7 1 3 7 1

产品:3 14 3 12 35 6 21 56 9

添加 $3 + 14 + 3 + 12 + 35 + 6 + 21 + 56 + 9$ 可得159，它不是10的倍数，因此原始数无效。
根据数字322 271 627:

3 2 2 2 7 1 6 2 7

乘以 : 3 7 1 3 7 1 3 7 1

产品:9 14 2 6 49 1 18 14 7

添加 $9 + 14 + 2 + 6 + 49 + 1 + 18 + 14 + 7$ 可得120，即10的倍数，因此原始数是有效的。

(尽管美联储的某些路由符号范围是保留的，因此没有分配，但验证算法不会检查保留的数字，以避免在ABA更改其政策时不得不修改它。)