

巧妙的标识符如何工作

目录

[简介](#)

[信用卡号](#)

[U.S. 社会保险编号](#)

[CUSIP编号](#)

[路由编号的ABA](#)

简介

本文描述巧妙的标识符，是内置的内容扫描模式检测数据特定类型。对于此版本，系统将实现信用卡号、美国社会保险编号、CUSIP路由编号的编号和ABA的巧妙的标识符。

在内部，一个巧妙的标识符与在某个方面检查候选匹配的验证功能一起包括匹配候选字符串的常规表示。例如，信用卡号的验证功能保证校验数字正确。

每个巧妙的标识符的常规表达将包括词间距锚点('b')在两端。例如，(这防止系统匹配美国社会保险编号在更加长的数字串中间。)为了简化，这些从如下解释省略。

因为常规表示找到的子链可能不验证，聪明的标识符实施一定小心关于重叠匹配。例如，过滤器为信用卡号扫描字符串4321 9999 9999 9995 1234 5678 9000应该看到有效信用卡号4321 9999 9999 9995的9999，即使可能的编号的一简单常规表示扫描将查找9999 4321 9999 9999和9995 1234 5678 9000。

信用卡号

信用卡号开始与一种可变长的卡类型，指示编号是否是签证，万事达卡，AMEX等等和与校验数字的末端。不同的卡类型在整个编号使用位不同的编号，但是校验数字计算是相同的在每个案件。

在途中注意，否则JCB卡没有匹配。并且，13位签证编号在我们的实施不存在和不会匹配。

16位信用卡号将匹配一以下常规表达：

```
[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}
```

当前缀是"4"，"51"-"55"或者"6011"。

15位AMEX编号将匹配一以下常规表达：

[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}

当允许的前缀是"34"或"37"。

14位Diners Club编号将匹配一以下常规表达：

[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}

当允许的前缀是"300"- "305"，"36"或者"38"。

注意常规表达定义了一特定分组一个给的信用卡长度的位，并且，如果有在位之间的标点，这必须始终是一样。

信用卡号的最终位是使用Luhn算法创建的校验数字。工作从编号的右端，请加倍每第二个数字。然后把发生的编号的单个数字(被加倍的两部分和不是)的那个加起来。如果结果是多个的10，则编号有效。

例如，给第1234 5678 9012 3456：

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

双：2 2 6 4 10 6 14 8 18 0 2 2 6 4 10 6

加2 + 2 + 6 + 4 + 1 + 0... + 1 + 0 + 6给64，不是多个的10，因此编号无效。

给第1234 5678 9876 3333：

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

双：2 2 6 4 10 6 14 8 18 8 14 6 6 3 6 3

加2 + 2 + 6 + 4 + 1 + 0... + 6 + 3给80，是多个的10，因此编号有效。

U.S. 社会保险编号

社会保险编号分开成一个3个位区域号，分配地理，在区域内的一个特定顺序分配的一个2个位组编号和顺序地分配的一4个位序列号。

我们的实施将使用以下常规表达：

[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}

这是上面表达式的一些示例：

555-55-5555
555.55.5555
555 55 5555

社会保险管理部门维护分配区域/组编号的列表：SSN发出的[3]。但是，因为本文周期地更改，我们不能取决于对此为验证。验证功能检查3个字段都不是所有零，并且前3个位少于800是。(上一个参考使用771作为限制，但是SSA已经有有前3个位的771和772指定号码。)

(开始与666的编号未分配，并且在范围987-65-4320至987-65-4329的编号为通告保留。并且，078-05-1120是被误用的SSN;它是秘书的实际SSN在钱包公司，使用编号为例。)

CUSIP编号

CUSIP (在统一安全识别步骤的委员会)编号是识别多种类型北美洲证券的9个字母数字标识符。编号分开成六字发布者编号，独特识别发布者(即，公司)，识别特定的安全的2个字符后缀;即，普通股，与优先股票与选项与固定收益仪器。

CUSIP聪明的标识符代码将使用以下常规表达：

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

验证功能类似于用于信用卡号的那个。唯一的差异是在CUSIP编号的字母转换对数值通过分配A=10，B=11，...，Z=35。

从cusip.com网站的一示例使用CUSIP第392690 QT 3：

3 9 2 6 9 0 问 T 3

转换字母：3 9 2 6 9 0 26 29 3

双：3 18 2 12 9 0 26 58 3

加3 + 1 + 8 + 2 + 1 + 2 + ... + 5 + 8 + 3给50，是多个的10，因此原号码有效。

路由编号的ABA

ABA (美国银行业协会)路由编号是9位值。前4个位是美联储路由符号的，下4机构标识符和为时校验数字。

ABA路由编号聪明的标识符代码将使用以下常规表达：

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

介入的验证功能乘每个位以3，7，1，...;如果产品的总和是多个的10，编号有效。

例如请认为第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

加 $3 + 14 + 3 + 12 + 35 + 6 + 21 + 56 + 9$ 给159, 不是多个的10, 因此原号码无效。

给第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

加 $9 + 14 + 2 + 6 + 49 + 1 + 18 + 14 + 7$ 给120, 是多个的10, 因此原号码有效。

(虽然美联储路由符号的某些范围保留, 并且不己分配, 验证算法不会检查保留编号, 避免必须修改它, 如果ABA更改其策略。)