

Funzionamento degli identificatori avanzati

Sommario

[Introduzione](#)

[Numeri carta di credito](#)

[Codici previdenziali USA](#)

[Numeri CUSIP](#)

[Numeri di routing ABA](#)

Introduzione

In questo documento vengono descritti gli identificatori intelligenti, ovvero modelli incorporati di scansione del contenuto che rilevano determinati tipi di dati. In questa release, il sistema implementerà gli identificatori intelligenti per i numeri di carta di credito, i numeri di previdenza sociale, i numeri CUSIP e i numeri di routing ABA.

Internamente, un identificatore intelligente è costituito da un'espressione regolare che corrisponde alle stringhe dei candidati, insieme a una funzione di convalida che controlla in qualche modo la corrispondenza dei candidati. Ad esempio, la funzione di convalida di un numero di carta di credito garantisce che la cifra di controllo sia corretta.

Le espressioni regolari per ogni identificatore smart includeranno ancoraggi limite di parola ('b') a entrambe le estremità. In questo modo si evita che il sistema corrisponda a un numero di previdenza sociale statunitense, ad esempio al centro di una stringa di cifre più lunga. Per semplicità, questi elementi vengono omessi dalle descrizioni riportate di seguito.

L'implementazione degli identificatori intelligenti deve prestare attenzione alle corrispondenze sovrapposte, poiché una sottostringa trovata dall'espressione regolare potrebbe non essere convalidata. Ad esempio, un filtro sta eseguendo la ricerca dei numeri di carta di credito in base alla stringa 9999 4321 9999 999 9995 1234 5678 9000 dovrebbe trovare il numero di carta di credito valido 4321 9999 999 9995, anche se una semplice ricerca di numeri tramite espressione regolare potrebbe trovare 999 4321 9999 999 e 99 995 1234 5678 9000.

Numeri carta di credito

Un numero di carta di credito inizia con un tipo di carta a lunghezza variabile, che indica se il numero è un VISA, una MasterCard, AMEX e così via, e termina con una cifra di controllo. Tipi di carta diversi utilizzano numeri di cifre diversi nell'intero numero, ma il calcolo della cifra di controllo è lo stesso in ogni caso.

Si noti che le schede enRoute o JCB non corrispondono. Inoltre, i numeri VISA a 13 cifre non esistono e non corrisponderanno nella nostra implementazione.

I numeri di carta di credito a 16 cifre corrispondono a una delle seguenti espressioni regolari:

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

Con il prefisso "4", "51"- "55" o "6011".

I numeri AMEX a 15 cifre corrispondono a una delle seguenti espressioni regolari:

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

Con i prefissi ammessi sono "34" o "37".

I numeri a 14 cifre di Diners Club corrispondono a una delle seguenti espressioni regolari:

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

Con i prefissi ammessi sono "300"- "305", "36" o "38".

Si noti che le espressioni regolari definiscono un raggruppamento specifico di cifre per una determinata lunghezza della carta di credito e che se vi è una punteggiatura tra le cifre, deve essere sempre la stessa.

La cifra finale in un numero di carta di credito è una cifra di controllo creata utilizzando l'algoritmo Luhn. A partire dall'estremità destra del numero, raddoppiare ogni due cifre. Sommare quindi le singole cifre dei numeri risultanti (sia quelle che sono state raddoppiate che quelle che non lo erano). Se il risultato è un multiplo di 10, il numero è valido.

Ad esempio, dato il numero 1234 5678 9012 3456:

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

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

Aggiungendo $2 + 2 + 6 + 4 + 1 + 0 \dots + 1 + 0 + 6$ si ottiene 64, che non è un multiplo di 10, quindi il numero non è valido.

Dato il numero 1234 5678 9876 3333:

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

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

Aggiungendo $2 + 2 + 6 + 4 + 1 + 0 \dots + 6 + 3$ si ottiene 80, che è un multiplo di 10, quindi il numero è valido.

Codici previdenziali USA

I numeri di previdenza sociale sono suddivisi in un numero di zona a 3 cifre, assegnato geograficamente, un numero di gruppo a 2 cifre assegnato in un particolare ordine all'interno di un'area e un numero seriale a 4 cifre assegnato in sequenza.

La nostra implementazione utilizzerà le seguenti espressioni regolari:

```
[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}
```

Di seguito sono riportati alcuni esempi delle espressioni riportate sopra:

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

Amministrazione previdenza sociale gestisce un elenco dei numeri di area/gruppo assegnati: SSN emesso [3]. Ma poiché questo documento cambia periodicamente, non possiamo fare affidamento su di esso per la convalida. La funzione di convalida verificherà che nessuno dei tre campi sia costituito da zero e che le prime tre cifre siano inferiori a 800. Il riferimento precedente utilizza 771 come limite, ma SSA ha già assegnato numeri con le prime tre cifre 771 e 772.

I numeri che iniziano con 666 non vengono assegnati e i numeri compresi tra 987-65-4320 e 987-65-4329 sono riservati alla pubblicità. Inoltre, 078-05-1120 è il SSN più utilizzato in modo improprio; è stato il numero SSN effettivo di una segretaria di una società di portafoglio a utilizzare il numero come esempio.)

Numeri CUSIP

I numeri CUSIP (Committee on Uniform Security Identification Procedures) sono 9 identificatori alfanumerici che identificano i titoli nordamericani di vario tipo. Il numero è diviso in un numero di 6 caratteri che identifica in modo univoco l'emittente (ad esempio una società), un suffisso di 2 caratteri che identifica il particolare titolo; ad esempio, azioni ordinarie rispetto alle azioni privilegiate rispetto alle opzioni rispetto agli strumenti a reddito fisso.

Il codice identificatore intelligente CUSIP utilizzerà le seguenti espressioni regolari:

```
[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 funzione di convalida è simile a quella utilizzata per i numeri di carta di credito. L'unica differenza è che le lettere nel numero CUSIP vengono convertite in un valore numerico assegnando A=10, B=11, ..., Z=35.

Un esempio tratto dal sito Web.cusip.com utilizza il numero CUSIP 392690 QT 3:

3 9 2 6 9 0 T 3

Converti lettere: 3 9 2 6 9 0 26 29 3

Doppio: 3 18 2 12 9 0 26 58 3

Aggiungendo $3 + 1 + 8 + 2 + 1 + 2 + \dots + 5 + 8 + 3$ si ottiene 50, che è un multiplo di 10, quindi il numero originale era valido.

Numeri di routing ABA

Un numero di routing ABA (American Banking Association) è un valore a 9 cifre. Le prime 4 cifre sono il simbolo dell'instradamento della Federal Reserve, le 4 successive il codice di identificazione dell'istituzione e l'ultima una cifra di controllo.

Il codice di identificazione intelligente del numero di routing ABA utilizzerà le seguenti espressioni regolari:

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

La funzione di convalida implicava il moltiplicare ogni cifra per 3, 7, 1, ...; se la somma dei prodotti è un multiplo di 10, il numero è valido.

Si consideri ad esempio il numero 123 456 789:

1 2 3 4 5 6 7 8 9

Moltiplica per: 3 7 1 3 7 1 3 7 1

Prodotto: 3 14 3 12 35 6 21 56 9

Aggiungendo $3 + 14 + 3 + 12 + 35 + 6 + 21 + 56 + 9$ si ottiene 159, che non è un multiplo di 10, pertanto il numero originale non era valido.

Dato il numero 322 271 627:

3 2 2 2 7 1 6 2 7

Moltiplica per: 3 7 1 3 7 1 3 7 1

Prodotto: 9 14 2 6 49 1 18 14 7

Aggiungendo $9 + 14 + 2 + 6 + 49 + 1 + 18 + 14 + 7$ si ottiene 120, che è un multiplo di 10, pertanto il numero originale era valido.

(Anche se alcuni intervalli di simboli di routing della Federal Reserve sono riservati, e quindi non assegnati, l'algoritmo di convalida non verificherà i numeri riservati, per evitare di doverli rivedere se l'ABA cambia politica.)