

Hoe slimme identificatoren werken

Inhoud

[Inleiding](#)

[Creditkaartnummers](#)

[Amerikaanse nummers voor sociale zekerheid](#)

[CUSIP-nummers](#)

[ABA-routingnummers](#)

Inleiding

Dit document beschrijft slimme herkenningstekens, die ingebouwde scanpatronen van content zijn die bepaalde soorten gegevens herkennen. Voor deze release zal het systeem slimme identificatoren implementeren voor creditcardnummers, US Social Security-nummers, CUSIP-nummers en ABA-routingnummers.

Intern bestaat een slimme identicator uit een reguliere expressie die overeenkomt met kandidaat strings, samen met een validatiefunctie die de kandidaat match op een of andere manier controleert. De validatiefunctie voor een creditcardnummer zorgt er bijvoorbeeld voor dat het controlecijfer juist is.

De reguliere expressies voor elke slimme identicator omvatten woord grensankors ('b') aan beide uiteinden. (Dit voorkomt dat het systeem een Amerikaans sociaal zekerheidsnummer overeenkomt, bijvoorbeeld in het midden van een langere reeks cijfers.) Voor de eenvoud worden deze weggelaten in de onderstaande beschrijvingen.

De slimme herkenningstoepassing moet voorzichtig zijn bij overlappende overeenkomsten, omdat een substring die gevonden wordt door de reguliere expressie mogelijk niet valideert. Zo wordt een filter bijvoorbeeld gescand voor creditcardnummers op de string 9999 4321 9999 9999 99995 1234 5678 9000 het geldige creditcardnummer 4321 9999999. 995, ook al zou een eenvoudige reguliere expressiescan voor mogelijke getallen 999 4321 9999 9999 en 9995 1234 5678 9000 vinden.

Creditkaartnummers

Een creditcardnummer begint met een variabel lengte kaarttype, wat aangeeft of het nummer een VISA, MasterCard, AMEX enz. is en eindigt met een controlecijfer. Verschillende kaarttypen gebruiken verschillende cijfers in het gehele nummer, maar de berekening van het controlecijfer is in elk geval hetzelfde.

Merk op dat nRoute- of JCB-kaarten niet overeenkomen. Bovendien bestaan er geen 13-cijferig VISA-nummers en worden deze niet aangepast in onze implementatie.

De 16-cijferige creditcardnummers komen overeen met een van de volgende reguliere expressies:

```
[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}
```

Als het voorvoegsel "4", "51"- "55" of "6011" is.

De 15-cijferig AMEX-getallen komen overeen met een van de volgende reguliere expressies:

```
[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}
```

De toegestane voorfixes zijn "34" of "37".

De 14-cijferige Diners Club getallen komen overeen met een van de volgende reguliere expressies:

```
[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}
```

De toegestane voorfixes zijn "300"- "305", "36" of "38".

Merk op dat de reguliere expressies een specifieke groep van cijfers definiëren voor een bepaalde lengte van de kredietkaart, en dat, indien er sprake is van punctuatie tussen de cijfers, dit overal hetzelfde moet zijn.

Het laatste cijfer in een creditcardnummer is een controlecijfer dat is gecreëerd met behulp van het Luhn-algoritme. Vanaf het rechteruiteinde van het nummer, dubbelvoudig elke tweede cijfer. Voeg vervolgens de individuele cijfers van de resulterende getallen toe (zowel de getallen die zijn verdubbeld als de getallen die niet fungeerden). Als het resultaat een veelvoud van 10 is, dan is het aantal geldig.

Bijvoorbeeld, gegeven het nummer 1234 5678 9012 3456:

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

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

Toevoegen 2 + 2 + 6 + 4 + 1 + 0 ... + 1 + 0 + 6 geeft 64, wat geen veelvoud van 10 is, zodat het nummer ongeldig is.

Gegeven nummer 1234 5678 9876 3333:

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

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

Toevoegen $2 + 2 + 6 + 4 + 1 + 0 \dots + 6 + 3$ geeft 80, wat een veelvoud van 10 is, zodat het nummer geldig is.

Amerikaanse nummers voor sociale zekerheid

De sociale - zekerheidsnummers worden verdeeld in een 3-cijferig gebiedsnummer, dat geografisch is toegewezen, een 2-cijferig groepsnummer dat in een bepaalde volgorde binnen een gebied wordt toegewezen, en een 4-cijferig serienummer dat achtereenvolgens is toegewezen.

Onze implementatie zal de volgende reguliere expressies gebruiken:

```
[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 zijn een paar voorbeelden van de expressies hierboven:

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

De Dienst voor sociale zekerheid houdt een lijst bij van de toegewezen gebieden/groepen: Uitgegeven SSN [3]. Maar omdat dit document van tijd tot tijd verandert, kunnen we er niet op vertrouwen voor validatie. De validatiefunctie zal controleren of geen van de drie velden alle nullen zijn en dat de eerste 3 cijfers minder dan 800 zijn. (De vorige referentie gebruikt 771 als limiet, maar de SSA heeft al getallen toegewezen met de eerste 3 cijfers 771 en 772.)

(Nummers die beginnen met 666 worden niet toegewezen, en getallen in de nummers 987-65-4320 tot en met 987-65-4329 zijn voorbehouden voor reclame. Ook is 078-05-1120 het meest gebruikte SSN; het was het SSN van een secretaresse van een beursgenoteerde onderneming, die het nummer als voorbeeld gebruikte.)

CUSIP-nummers

CUSIP (Committee on Uniform Security Identification Procedures) zijn 9 alfanumerieke identificatienummers die Noord - Amerikaanse effecten van verschillende types identificeren. Het aantal wordt opgesplitst in een uitgevende instelling van 6 tekens, die de uitgevende instelling (bijvoorbeeld een onderneming) uniek identificeert, en een achtervoegsel van 2 tekens dat het specifieke effect identificeert; Bijvoorbeeld gewone aandelen, vs. preferente aandelen vs. optie vs. instrument met vaste inkomens.

De slimme identificatiecode van CUSIP zal de volgende reguliere expressies gebruiken:

```
[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]
```

De validatiefunctie is vergelijkbaar met die welke gebruikt wordt voor creditcardnummers. Het enige verschil is dat letters in het CUSIP-nummer worden geconverteerd naar een numerieke

waarde door A=10 toe te wijzen, B=11, ..., Z=35.

Een voorbeeld van de website cusip.com gebruikt het CUSIP-nummer 392690 QT 3:

3 9 2 6 9 0 KW T 3

Brieven converteren: 3 9 2 6 9 0 26 29 3

Dubbel: 3 18 2 12 9 0 26 58 3

Toevoegen $3 + 1 + 8 + 2 + 1 + 2 + \dots + 5 + 8 + 3$ geeft 50, wat een veelvoud van 10 is, zodat het oorspronkelijke nummer geldig was.

ABA-routingnummers

Een ABA (American Banking Association)-routingnummer is een 9-cijferige waarde. De eerste 4 cijfers zijn het symbool van de Federal Reserve, het volgende 4 de instelling identifier en het laatste controlecijfer.

De ABA-routingnummer slimme identificatiecode gebruikt de volgende reguliere expressies:

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

De valideringsfunctie die erin bestaat elk cijfer te vermenigvuldigen met 3, 7, 1, ...; indien de som van de producten meer dan 10 bedraagt, is het nummer geldig.

Bijvoorbeeld nummer 123 456 789:

1 2 3 4 5 6 7 8 9

Vermenigvuldigen met: 3 7 1 3 7 1 3 7 1

Product: 3 14 3 12 35 6 21 56 9

Toevoegen $3 + 14 + 3 + 12 + 35 + 6 + 21 + 56 + 9$ geeft 159, wat geen veelvoud van 10 is, zodat het oorspronkelijke nummer ongeldig was.

Gegeven nummer 322 271 627:

3 2 2 7 1 6 2 7

Vermenigvuldigen met: 3 7 1 3 7 1 3 7 1

Product: 9 14 2 6 49 1 18 14 7

Toevoegen $9 + 14 + 2 + 6 + 49 + 1 + 18 + 14 + 7$ geeft 120, wat een veelvoud van 10 is, zodat het oorspronkelijke nummer geldig was.

(Hoewel bepaalde marges van de door de Federal Reserve toegekende routingsymbolen gereserveerd zijn en daarom niet zijn toegewezen, zal het validatiealgoritme niet voor gereserveerde getallen controleren, om te vermijden dat het moet worden herzien als het ABA zijn beleid wijzigt.)