Nog meer voorbeelden

Even terugkijken.....wat hebben we tot nu toe gedaan.

Tot nu toe hebben we teksten gevalideerd. We hebben gekeken of een string voldoet aan een bepaald patroon.

We hebben bijvoorbeeld bepaald of een getal een valide 06-nummer is, hiervoor gebruikte we de expressie:

`^06\d{8}$`	begin met 06 en dan precies 8 decimalen de d staat voor decimalen en {8} staat voor 8 posities

We hebben verder bepaald of een string een hex getal is met de reguliere expressie:

`^[0-9A-F]+$`	begin met een 0,1,2,3,4,5,6,7,8,9,0,A,B,C,D,E,F en herhaal dat 1 of meer keren tot het eind.

Als we willen weten of er ergens in de tekst een 06-nummer staat dan kunnen we de ^ en $ weglaten. Er wordt dan niet meer van het begin naar het einde gezocht, maar gekeken of er ergens een rij karakters is dat voldoet aan de expressie.

In de laatste reguliere expressie zie je een plus. In reguliere expressies hebben de symbolen `?`, `+` en `*` verschillende betekenissen en worden ze gebruikt om het aantal herhalingen van een patroon in de tekst te bepalen.

Teken	**Betekenis**	**Voorbeeld**	
	als je niets plaats dan moet het element 1x voorkomen.	A	Alleen A voldoet
+	een plus betekent 1 of meer keren	AB+	AB en ABBB voldoen
*	een * betekent 0, 1 of meer keren	AB*C	AC en ABC of ABBBC voldoen
?	een ? betekent 0 of 1 keer	AB?C	AC en ABC voldoen

Voor het teken kan des een teken staan of een groep tekens, je kunt ook een . Gebruiken.

ABC+	A komt 1 keer voor dan een B en dan één of meer keren een C. *ABCCC* is dus goed, *AABC* niet.
[A-C]+	A,B of C komen 1 of meer keren voor. *ABCAABBAA* is dus goed en A ook.

[A-C]{5,6}	A,B, of C en dat minimaal 5 en maximaal 6 keer. *CBBCBA* is dus goed, maar *CBBCBAC* niet omdat dat een karakter te veel is.
.*	Elk karakter (any) komt 0 of meer keren voor (alles matched dus).

En dan kun je nog ^ of $ gebruiken.

Expressie	Uitleg	Voorbeeld van een match
^ABC	De tekst begint met ABC, daarna mag alles	ABC zijn de beginletters van het alfabet.
^AB*C	De tekst begint met een A dan 0,1, of meer B's en dan een C, daarna mag alles	AC wordt gebruikt als afkorting voor airconditioner
^A*B*	De tekst mag alleen 0,1 of meer A;s bevatten gevolgd door 0,1 of meer B's.	BB staat for bed & breakfast
^[A-Z]+$	De tekst bestaat alleen uit hoofdletters en er moet er minimaal één voorkomen.	KIJK UIT

De validatie in python doe je met het commando:

```
re.match(expressie, text)
```

Opdracht

Jij krijgt nu een tabel met een expressie. De kolom uitleg en match is leeg. Jij moet die kolommen vullen. Leg uit wat de expressie doet en bepaal of de tekst een match is. Probeer de expressie uit met de code van de vorige opgave.

Expressie	Tekst	Match?	Uitleg in eigen woorden (waarom wel of geen match)
expressie= r'[0-9]+^'	Mijn nummer is: 06 - 1234 1234		
expressie= r'[0-9]{8,10}^'	Mijn nummer is: 06 - 1234 1234		
expressie= r'$[A-Z]+'	Mijn nummer is: 06 - 1234 1234		
expressie= r'max'	Maximale temperatuur is 12 graden		
expressie= r'[max]'	x-men		

expressie= r'[k-mK-M]'	Dit is een voorbeeld		
expressie= r'$[k-mK-M]'	Matig je alcohol gebruik!		
expressie= r'$[k-mK-M]'	Gaat de KLM failliet?		
expressie= r'^[a-kA-K]+$'	De kakkerlak kakt		
expressie= r'$[k-mK-M]$'	De kakkerlak kakt erg veel		

Test alle reguliere expressies en zet in de kolom match of het voorbeeld een matched met de reguliere expressie.

Gebruik de code van de vorige opdracht om je antwoorden te controleren.

Inleveren

1. PDF met jouw naam en de tabel waarin je in de laatste kolom met *true* of *false* (wel of geen match) aangeeft en daarachter een uitleg geeft.
2. De code waarmee je de voorbeelden hebt gecontroleerd.

--

Revision #3
Created 2023-07-22 07:49:31 UTC by Max
Updated 2023-07-31 10:51:32 UTC by Max