

Aufgaben Übung Reguläre Ausdrücke

Zusammenfassung reguläre Ausdrücke

a	Das Zeichen a
.	Ein beliebiges Zeichen
[abc]	Ein beliebiges Zeichen aus der Menge {a, b, c}
\d	eine Ziffer [0-9]
\w	Buchstabe, Ziffer oder Unterstrich – [a-zA-Z_0-9]
X Y	X oder Y
X*,X*?	Eine beliebige Wiederholung von X, gierig bzw. genügsam
X+, X+?	Mindestens einmal X, gierig bzw. genügsam
X?	Höchstens einmal X
{n}	der vorangehende Ausdruck muss exakt n mal vorkommen
{m,n}	der vorangehende Ausdruck muss mindestens m mal vorkommen und darf höchstens n mal vorkommen
(xyz)	Gruppierung: xyz müssen miteinander vorkommen
[^xyz]	Nicht x, nicht y, nicht z
\X	X ein Sonderzeichen \()[\]*+?{} ^\$-

Aufgabe: Datumserkennung anpassen

In einem Programm, das Webseiten auf Daten hin durchsucht, finden Sie folgenden regulären Ausdruck: `\d\d\.\d\d\.\d\d`

Beschreiben Sie in Worten, (1) welche Daten dieser Ausdruck findet und (2) welche Daten er nicht findet:

Geben Sie Beispiele für alle Arten von Daten an, die gemäss folgender Beschreibung zulässig sind:

- sowohl zweistellige als auch vierstellige Jahreszahlen sind zulässig
- sowohl einstellige als auch zweistellige Tag- und Monatsangaben sind zulässig

Erweitern Sie den Ausdruck `\d\d\.\d\d\.\d\d`, so dass er Daten gemäss folgender Beschreibung erkennt:

- sowohl zweistellige als auch vierstellige Jahreszahlen sind zulässig
- sowohl einstellige als auch zweistellige Tag- und Monatsangaben sind zulässig

Aufgabe: Postleitzahlen erkennen

Postleitzahlen in der Schweiz sind vierstellig im Bereich von 1000..9999.

Schreiben Sie einen regulären Ausdruck, der Postleitzahlen erkennt:

Aufgabe: Zahlen mit Tausender-Apostroph erkennen

Zahlen gibt es in den unterschiedlichsten Formaten. Betrachten wir folgende Zahlen:

- * nur positive ganze Zahlen
- * Tausender-Apostroph zur Abtrennung von 1000 Bereichen
- * zum Beispiel 1, 42, 486, 9'386, 719'528, 783'748'894'846

Auch hier hat jemand von einem Anfang gemacht: Der Ausdruck `\d{1,3}'\d{3}` erkennt immerhin schon die Zahlen 1'000 bis 999'999.

Erweitern Sie den Ausdruck, so dass die Zahlen wie eingangs beschrieben erkennt:

Aufgabe: Mailadressen erkennen

Sie finden in einem Programm folgenden regulären Ausdruck für die Überprüfung, ob eine Eingabe eine gültige Mailadresse darstellt:

```
[a-z][A-Z][0-9]_|\.)+@[a-z][A-Z][0-9]_|\.)+\.([a-z][A-Z][0-9]_|\.)+
```

Beschreiben Sie in Worten, was dieser Ausdruck erkennt:

Bei Mailadressen kommt nach dem letzten Punkt eine sog. Toplevel Domain wie .ch oder .com.

Geben Sie vier unterschiedliche Beispiele von Mailadressen an, die der obige Ausdruck akzeptiert, die aber keine gültige Toplevel Domain enthalten:

Aufgabe: Titel in HTML-Seiten suchen

Webseiten werden textuell beschrieben. Eine einfache Webseite könnte sein:

```
<h1>Dies ist eine Überschrift ersten Grades</h1>
```

```
<h2>Dies ist eine Überschrift zweiten Grades</h2>
```

```
<p>Dies ist einfach nur ein ganz normaler Abschnitt.</p>
```

```
<h1>Dies ist noch eine Überschrift ersten Grades</h1>
```

Die Befehle in spitzen Klammern geben dabei den Browsern an, welche Bedeutung welcher Abschnitt einer Webseite hat.

Schreiben Sie einen regulären Ausdruck, der Überschriften ersten Grades erkennt.

Schreiben Sie einen regulären Ausdruck plus einen ersetzenden Ausdruck, der unterstrichene Texte (<u>...</u>) durch kursive Texte (<i>...</i>) ersetzt.