

Quickstart – Reguläre Ausdrücke

Reguläre Ausdrücke sind Muster zur Beschreibung von Mengen (bzw. Untermengen) von Zeichenketten mithilfe von **Metazeichen** und einer vorgegebenen Syntax. Sie dienen der Optimierung jeglicher zeichenbasierter Operationen bei der Verarbeitung von Zeichenketten (in erster Linie Suchoperationen). Hierdurch wird eine einzelne Auflistung möglicher Alternativen umgangen. Zum Beispiel:

`\b[Vv]o[nm]\s+([A-Za-zÄÖÜäöüß]+\s+){1,10}aus\b1`

A. Zeichen und Zeichenklassen

<code>a</code>	matcht das Zeichen a (case sensitive!)
<code>.</code>	matcht ein beliebiges Zeichen
<code>[ui]</code>	Zeichenklasse matcht hier das Zeichen „u“ oder „i“
<code>[^AaEeIiOoUu]</code>	negierte Zeichenklasse matcht hier alle Zeichen außer Vokale
<code>[A-Ka-k] [0-5]</code>	Schreibweise für Zeichenklassen mit einem Bereich von Zeichen

Beispiele:

`B[ui]rma` → matcht „Burma“ sowie „Birma“

`M[ae][iy]er` → matcht verschiedene Schreibweisen eines Namens

B. Vordefinierte Zeichenklassen (Auswahl)

<code>\d</code>	eine Ziffer
<code>\D</code>	ein Zeichen, das keine Ziffer ist
<code>\w</code>	ein Wortzeichen, Ziffer oder Unterstrich (jedoch keine Umlaute und „ß“)
<code>\W</code>	ein Zeichen, das weder Buchstabe noch Zahl noch Unterstrich ist
<code>\s</code>	Whitespace
<code>\S</code>	ein Zeichen, das kein Whitespace ist

Wichtig!!!

`\w` → matcht Wortzeichen, i. d. R jedoch keine Umlaute und „ß“

Abhilfe: **`[A-Za-zÄÖÜäöüß]`** für Wortzeichen

C. Quantoren

<code>?</code>	Ausdruck einmal erlaubt, aber optional
<code>*</code>	jede Anzahl erlaubt, auch 0
<code>+</code>	Mindestens einmal, mehr erlaubt
<code>{n}</code>	Festgelegte Anzahl <i>n</i>
<code>{min,max}</code>	<i>mindestens</i> gefordert, <i>maximal</i> erlaubt

¹ Entspricht folgender Suchanfrage: Von ... aus / Vom ... aus / von ... aus / vom ... aus

Beispiele:

[A-K]\w+ → matcht Wörter aus mindestens zwei Buchstaben, die mit großem „A“ bis „K“ anfangen

\d{1,2}\.\d{1,2}\.\d{2,4} → matcht Datumsangaben

D. Sonstige Metazeichen (Auswahl)

	passt auf mindestens einen der alternativen Ausdrücke
(... ...)	Beschränkt den Geltungsbereich von Alternationen
\b	matcht Wortgrenzen
\	escape (Aufhebung von Metazeichen)

Beispiele:

\. → matcht einen Punkt

[Dd](er|ie|as) → matcht bestimmte Artikel

[a-z0-9-_.]+@[a-z0-9-_.]+\.\w{2,3} → matcht E-Mail Adressen

Durch die Kombination von regulären Ausdrücken auf verschiedene Weise, können zwar komplexe Suchanfragen formuliert werden, doch ist ihr Gebrauch nicht immer ganz einfach. Wir raten bei der Arbeit mit EXAKT, Dokumentationen (Literatur/ Websites) zu *Regulären Ausdrücken* hinzuzuziehen.



Technische Einzelheiten (für Nutzer mit Vorkenntnissen) zur exakten Syntax und zu ihrem Gebrauch können über den folgenden Link aufgerufen werden:

<http://java.sun.com/javase/6/docs/api/java/util/regex/Pattern.html>