

RegEx Cheat Sheet

Eine ganz kurze Einführung in reguläre Ausdrücke
(ohne Anspruch auf Vollständigkeit)
insbesondere für Java

Version: 0.4 (draft)
Datum: 08.12.2007
Autor: [Sascha Leib](#)

RegEx sind formalisierte Zeichenmuster, welche für Vergleiche von Zeichenketten benutzt werden können. Diese werden in Java u.a. von einer Reihe von Funktionen (z.B. „`matches()`“) verwendet.

In den meisten Fällen kann ein String durch sich selbst dargestellt werden. z.B.:

String	Regex	match
Hallo	Hallo	✓

Spezielle Symbole werden durch einen vorangestellten *Backslash* spezifiziert:

Darstellung	Beschreibung
\\	Das <i>Backslash</i> -Zeichen selbst (,\\')
\\t	Tabulator (,\\u0009')
\\n	Zeilenvorschub (,newline' = ,\\u000A')
\\r	Wagenrücklauf (,carriage-return' = ,\\u000D')
\\f	Seitenvorschub (,form feed' = ,\\u000C')
\\a	Signalzeichen (,alert' oder ,bell' = ,\\u0007')
\\e	Escape-Symbol (,\\u001B')
\\xhh bzw. \\xhhhh	Das Zeichen mit dem Hexadezimalwert hh bzw. hhhh
\\0n bzw. \\0nn bzw. \\0nnn	Das Zeichen mit dem Oktalwert n bzw. nn bzw. nnn

Achtung: im Java-Quelltext muss der *Backslash* selbst auch kodiert werden, für den doppelten Backslash schreibt dann also in den meisten Fällen „\\\\“.

Die folgenden Symbole stehen nicht für Zeichen sondern betreffen Begrenzungen im Zieltext:

^	Zeilenanfang (außerhalb einer Zeichenauswahl, s.u.)
\$	Zeilenende.
\\b	Wortgrenze.
\\B	Nicht-Wortgrenze.
\\A	Anfang des Eingabetextes.
\\G	Ende der letzten Fundstelle (bei Mehrfachsuche)
\\Z	Ende des Eingabetextes, ausgenommen Terminator.
\\z	Ende des Eingabetextes.

Daneben kennt RegEx Platzhalter („wildcards“) für beliebige Zeichen:

.	Genau ein beliebiges Zeichen.
*	Null oder mehr beliebige Zeichen.

Sollen diese Platzhalterzeichen selbst gesucht werden, müssen sie mit einem Backslash *escaped* werden, also als „\\.“ und „*“.

Zeichenklassen

Eine Auswahl von möglichen Zeichen kann in eckigen Klammern angegeben werden:

[abc]	Einer der Kleinbuchstaben ‚a‘, ‚b‘ oder ‚c‘.
[a-z]	Ein beliebiger Kleinbuchstabe (von ‚a‘ bis ‚z‘)
[a-zA-Z]	Ein beliebiger Groß- oder Kleinbuchstabe (‚a‘ bis ‚z‘ oder ‚A‘ bis ‚Z‘).
[0-9]	Eine beliebige Ziffer (‚0‘ bis ‚9‘)

Das Carret-Zeichen (‚^‘) kehrt die Auswahl um (*negation*). Dafür muss das Zeichen direkt nach der Klammer erscheinen, andernfalls wird es als Literal angenommen:

[^abc]	Ein beliebiges Zeichen, <i>außer</i> einem der Kleinbuchstaben ‚a‘, ‚b‘ oder ‚c‘.
[a-z^]	Ein beliebiger Kleinbuchstabe <i>oder</i> das Carret-Zeichen (!)

Zeichenklassen können miteinander verbunden werden:

[a-zA-Z]	Vereinigung: ‚a‘ bis ‚z‘ oder ‚A‘ bis ‚Z‘ (entspricht ‚[a-zA-Z]‘)
[a-m&&[l-z]]	Schnittmenge aus ‚a-m‘ und ‚l-z‘ (entspricht ‚[l-m]‘)
[a-z&&[^l-m]]	Ein Zeichen aus ‚a-z‘, ohne den Bereich ‚l-m‘ (entspricht ‚[a-k[n-z]]‘)

Einige wichtige Zeichenbereiche sind als Shortcuts vordefiniert:

\d	eine beliebige Ziffer (entspricht ‚[0-9]‘)
\D	Ein beliebiges Zeichen, ausgenommen Ziffern (entspricht ‚[^0-9]‘)
\s	Ein beliebiger Leerraum (entspricht ‚[\t\n\r\f]‘)
\S	Ein beliebiges Zeichen, ausgenommen Leerraum (Umkehrung von ‚\s‘)
\w	Textzeichen (entspricht ‚[a-zA-Z_0-9]‘)
\W	Nicht-Textzeichen (entspricht ‚[^a-zA-Z_0-9]‘)

Wiederholungen

Durch Anhängen dieser Suffixe kann man deklarieren, wie oft ein Muster vorhanden sein muss:

$X?$	Null oder ein Vorkommen von X
X^*	Null oder beliebig viele Vorkommen von X
X^+	Ein oder beliebig viele Vorkommen von X
$X\{n\}$	Genau n Vorkommen von X
$X\{n, \}$	Mindestens n Vorkommen von X
$X\{n, m\}$	Mindestens n und höchstens m Vorkommen von X

Auswahl

Außerhalb von Symbolklassen können Auswahlen wie folgt definiert werden:

XY	Muster X gefolgt von Muster Y
$X Y$	Muster X oder Muster Y

Gruppen

(X)	Definiert eine Gruppe
\n	entspricht dem Inhalt der n-ten (zuvor definierten) Gruppe (Ziffer)

Solche Gruppierungen sind v.a. interessant für Ersetzungsvorgänge:

String	Suchmuster	Ersetzungsmuster	Ergebnis
Peter Panter	(\w*) (\w*)	\2, \1	Panter, Peter
Peter Panter	([A-Z])\w* (\w*)	\1. \2	P. Panter

Hinweise:

- Die Gruppe Null („\0“) entspricht immer dem Text, der vom gesamten Suchmuster gefunden wurde;
- Verschachtelte Gruppen werden von außen nach innen durchnummeriert;
- Gruppen, die mit einem Fragezeichen beginnen, haben eine besondere Bedeutung, und werden bei der Zählung nicht berücksichtigt.

