

Regex

Vortrag von Alena Endres
JGW-1 Information Retrieval
Papenburg 2016

Was ist das?

- Regex (oder RegExp) = regular Expression
- Bestimmte syntaktische Regeln
- Benötigt für Suchen in Strings:
 - Muster des Strings wird mit dem vorgegebenen Regulären Ausdruck verglichen
- Zu jedem regulären Ausdruck gibt es eine zugehörige reguläre Sprache

Definition

Sei A ein Alphabet.

Eine Sprache $L (A^*)$ heißt regulär, wenn sie sich durch endlich viele Anwendungen der Operationen Vereinigung und Verkettung erzeugen lässt.

Der entsprechende Ausdruck, der angibt, in welcher Reihenfolge welche Operationen auf die Sprachen angewendet wird, heißt regulärer Ausdruck.

Quantifier (Wiederholungen)

Zeichen	Bedeutung
$a?$	a kommt kein- oder einmal vor.
a^+	a kommt mindestens einmal vor.
a^*	a kommt kein oder mehrmals vor.
$a\{3\}$	a kommt genau drei mal vor.
$a\{3,\}$	a kommt mindestens drei mal vor.
$a\{,5\}$	a kommt maximal fünf mal vor.
$a\{3,5\}$	a kommt mindestens drei, aber maximal fünf mal vor.
a	Die Zeichenkette beginnt mit einem Text der auf a passt
$ausdruck\$$	Die Zeichenkette endet mit einem Text der auf $ausdruck$ passt
$^ab\$$	Die Zeichenkette passt von Anfang bis Ende auf ab



Sollen mehrere verschiedene Zeichen angesprochen werden, muss um diese eine runde Klammer gesetzt werden: $(ab)?$ würde folgendes bedeuten: ab kommt ein- oder keinmal vor.

Zeichenklassen

Zeichen	Bedeutung
.	Jegliches Zeichen (Ausnahme der Zeilenumbruch, siehe auch bei den Modifiern)
A-Z	ABCDEFGHIJKLMNOPQRSTUVWXYZ
a-z	abcdefghijklmnopqrstuvwxyz
0-9 oder \d	0123456789
\w	A-Za-z0-9
\s	s.g. Whitespaces, also Leerzeichen, Zeilenumbrüche, Tabulatoren etc.

Weitere Zeichen

Symbol	Bedeutung
\	Vor Sonderzeichen
	Oder

Beispiele

$\Sigma = (1, 2, 3, 4, 5, 6, 7, 8, 9)$

a) Eine unendlich lange Zahl (Integer)

→ $[0-9]^+$

b) Eine unendlich lange Dezimalzahl (Float)

→ $[0-9]^+ (, [0-9]^+)$

c) Eine positive oder negative Dezimalzahl

→ $[+-]? [0-9]^+ (, [0-9]^+)$

Quellen

- <http://www.iti.fh-flensburg.de/lang/theor/regulaerer-ausdruck.htm>
- <http://www.webmasterpro.de/coding/article/regex-syntax.html>
-