

## Endliche Automaten und reguläre Ausdrücke



### Hinweis

Deterministische Endliche Automaten (DEA) lassen sich in reguläre Ausdrücke umwandeln und umgekehrt!



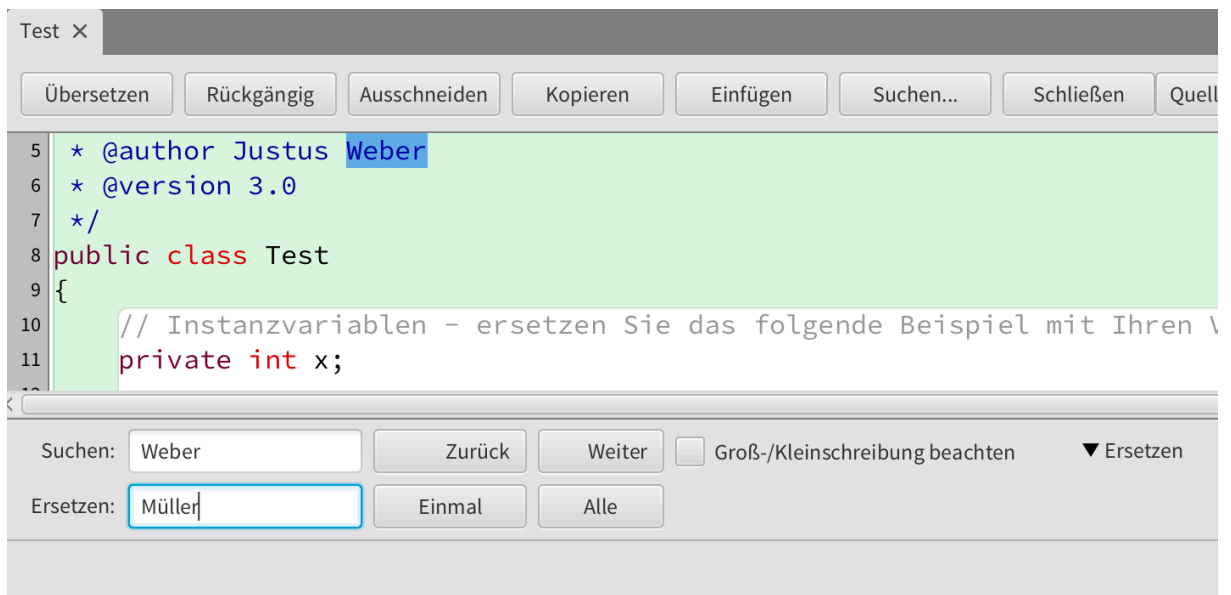
### Achtung

Die Definitionen und Aufgaben stammen zu einem großen Teil von der Internetseite FLACI und sind keine Erfindung von mir!



### Regulärer Ausdruck

Mit regulären Ausdrücken lassen sich reguläre Sprachen besonders kompakt und intuitiv beschreiben. Reguläre Ausdrücke werden beispielsweise in Texteditoren oder auf der Kommandozeile zum Suchen und Filtern verwendet. Sie werden auch zur Validierung von Benutzereingaben in (Web-)Anwendungen eingesetzt.



### Schreibweise (1):

Reguläre Ausdrücke lassen sich verketteten:  $a$  gefolgt von  $b$ , gefolgt von  $c$  wird kurz  $abc$  geschrieben.

#### Beispiel:

abacab



### Schreibweise (2):

Mit einem **senkrechten Strich |** werden Alternativen (sprich: „oder“) zwischen zwei oder mehreren regulären Ausdrücken notiert.

**Katze | Maus** beschreibt die Sprache  $L = \{\text{Katze}, \text{Maus}\}$ . Mehrere Alternativen werden einfach hintereinander geschrieben: **aaa | bbb | ccc**.

Deren Notationsreihenfolge spielt keine Rolle **aaa | bbb = bbb | aaa**.

Wichtig: Die Operation Verkettung hat Vorrang und wird zuerst ausgeführt. Möchten wir dies ändern, müssen wir runde Klammern einsetzen: **(Ha | Ma)us**.

- 1 Beschreibe einen regulären Ausdruck, der die Sprache der Farbwörter, die man mit den regulären Ausdrücken rot, gelb, grün und blau beschreibt, akzeptiert. Zeichne das zugehörige Syntaxdiagramm.

**Schreibweise (3):**

Mit ***a?*** kann ein Ausdruck ***a*** als optional gekennzeichnet werden. Er darf in einem passenden Wort höchstens einmal, d.h. einmal oder keinmal, vorkommen. Achtung: Diese Operation hat Vorrang vor der Verkettung. Der Ausdruck ***Haus?*** beschreibt somit die Sprache  $L = \{Hau, Haus\}$ , da das ***s*** als optional gekennzeichnet ist. Durch Klammerung lässt sich die Operation auf komplexere Teilausdrücke anwenden: ***das (blaue | rote )?Tuch***.

- 2 Beschreibe einen regulären Ausdruck, der alle drei Sätze mit ***Gelb*** beschreibt, aber nicht Sätze mit ***Rot?*** (Leerzeichen beachten!) Zeichne das zugehörige Syntaxdiagramm.
- Gelb ist eine tolle Farbe  
Gelb ist eine Farbe  
Gelbe Farbe  
Rote ist auch eine Farbe  
Rote Farbe

**Schreibweise (4):**

$a^*$  bedeutet, dass ein Ausdruck  $a$  beliebig oft (0-mal, 1-mal, 2-mal, ...) unmittelbar aufeinander folgen kann:  $\epsilon, a, aa, \dots$

Soll mindestens ein Zeichen  $a$  im Wort enthalten sein, kann statt  $aa^*$  die Kurzform  $a^+$  verwendet werden.

Achtung: Diese Operationen haben Vorrang vor der Verkettung. Der Ausdruck  $Haus^*$  beschreibt somit die Sprache  $L = \{Hau, Haus, Hauss, Hausss, \dots\}$ .

Runde Klammern beziehen diese Operation auf komplexere Teilausdrücke. Der reguläre Ausdruck  $(abc)^+$  beschreibt die Sprache der Form  $abc, abcabc, abcabcabc$  usw.

3

Verkürze den Ausdruck  $Haa^*ll^*oo^*$  mit der  $a^+$ -Schreibweise. Zeichne das zugehörige Syntaxdiagramm.

**Schreibweise (5):**

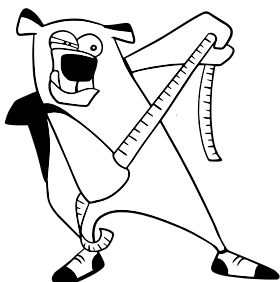
Der Ausdruck  $[abc]$  ist eine Zeichenklasse und beschreibt eine Menge, aus der genau ein Zeichen beliebig ausgewählt wird.

In einer Menge kann jedes einzelne Zeichen nur einmal angegeben werden. Die Reihenfolge spielt keine Rolle:  $[abc] = [cab]$ .

**Achtung:**  $[Handy]$  würde nicht etwa das Wort Handy erkennen, sondern genau einen einzelnen Buchstaben, nämlich entweder  $H, a, n, d$  oder  $y$ .

4

Durch Verkettung lassen sich mehrere reguläre Ausdrücke aneinanderfügen. Im Beispiel  $[rot][rot][rot]$  wird somit eine Folge von drei Zeichen gesucht. Wie müssen diese Zeichenmengen erweitert werden, damit die zulässigen Wörter auch die Grossbuchstaben R, O und T enthalten dürfen? Gib einen regulären Ausdruck an, der die Wörter **Rot, rot, Tor** und **tor** beschreibt. Zeichne das zugehörige Syntaxdiagramm.



**Schreibweise (6):**

In Zeichenklassen, wie z. B. `[a-z]`, `[a-zA-F]` oder `[0-9]`, können ganze Bereiche von Zeichen angegeben werden.

Das Minus an dieser Position hat damit die Bedeutung einer Aufzählung und muss ggf. am Ende oder Anfang genannt werden, wenn es selbst zur Sprache gehört `[abc-]`.

**5**

Beschreibe einen regulären Ausdruck, der erkennt, ob ein vierstelliger Zahlencode eingegeben wurde.

**Rechenweg**

Die wichtigsten Schreibweisen für reguläre Ausdrücke sind nun bekannt. Es gibt jedoch noch viele weitere Operationen, die hier nicht betrachtet werden.

Sucht man Sonderzeichen, wie `*`, `+`, `?`, `[]`, `()`, `$`, im Wort, so muss ein `\` vorangestellt werden: `5\+3`, `50\$` usw., um deren spezielle Funktion auszuschalten.

**6**

Beschreibe einen regulären Ausdruck, der erkennt, ob es sich um einen HTML-Farbcode handelt.

(Es sollen nur 6-stellige Farbcodes akzeptiert werden, die mit einem `#` beginnen. Es gibt nur Großbuchstaben oder Zahlen.)

**7**

Beschreibe einen regulären Ausdruck, der erkennt, ob es sich um eine gültige Uhrzeit handelt.

(*Beispiele:* `12:32`; `23:59`; `0:00`; `26:12`; `6:70`)

**8**

Beschreibe einen regulären Ausdruck, der erkennt, ob es sich um eine syntaktisch gültige Mailadresse handelt.

(*Beispiele:* `weber.justus@johann-2019.lernsax.de`; `test-schule@informatik.de`; `jw83opal@studserv.uni-leipzig.de`; `12ab@test.de`; `weber.justus@informatik`)