

Scanner mit regulären Ausdrücken

- JFlex generiert eine Klasse mit einer Methode, die beim Aufruf das nächste Symbol der Eingabe erkennt.
- Jedes Symbol wird durch einen regulären Ausdruck beschrieben.
- Dem Symbol ist ein Lexem zugeordnet (Folge der erkannten Zeichen).
- Der Endzustand mit dem längstmöglichen Lexem wird akzeptiert.
- Bei gleich langen Lexemen wird der Endzustand mit dem ersten Ausdruck akzeptiert.
- Dem Endzustand ist eine Aktion zugeordnet (in der Regel: `return Symbol`).
- Die erkannten Symbole gelten als gelesen.
- Reguläre Ausdrücke können auf bestimmte Zustände beschränkt sein.

JFlex erlaubt Abkürzungen durch reguläre Definitionen und Zeichenklassen.

Generierung

- Gleich zu behandelnde Zeichen werden zu Zeichenklassen zusammengefasst
- Umwandlung reguläre Ausdrücke \rightarrow NFA
- Umwandlung NFA \rightarrow DFA
- Umwandlung DFA \rightarrow minimaler DFA
- Speicherung in komprimierter Tabelle oder switch-Anweisung

NFA

Laufzeit: $O(n \times r)$

Speicherplatz: $O(n)$

DFA:

Laufzeit: $O(n)$

Speicherplatz: $O(2^r)$ (in der Regel aber nur $O(r)$)

(Grund: Zustände des DFA sind Teilmengen der Zustände des NFA)

Anmerkung: Java RegExp-Klasse hat ein anderes Ziel (Suche)