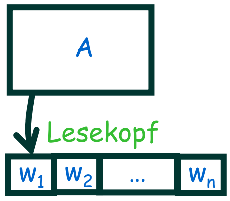
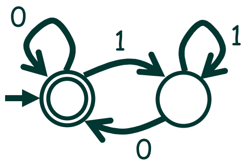
Endliche Automaten

Idee: endlicher Automat *A*

- hat endlich viele innere Zustände
- liest Eingabe $w \in \Sigma^*$ zeichenweise von links nach rechts
- gibt zum Schluß eine Ja/Nein Antwort



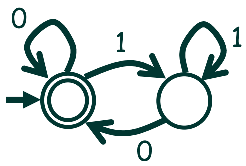
gelesenes Symbol w_i + aktueller innerer Zustand
→ neuer innerer Zustand



Deterministische endliche Automaten

Def.: Ein deterministischer endlicher Automat *A* (dea) ist ein 5-Tupel (Q,Σ,δ,q_0,F) mit

- Q endliche Menge von inneren Zuständen
- Σ endliches Alphabet
- $\delta : Q \times \Sigma \rightarrow Q$ Zustandsüberführungsfunktion
- $q_0 \in Q$ Anfangszustand
- $F \subset Q$ Menge der akzeptierenden Zustände



k-Schritt Überführungsfunktion

Def.: Sei $A=(Q,\Sigma,\delta,q_0,F)$ dea. Für $k\geq 1$ heißt

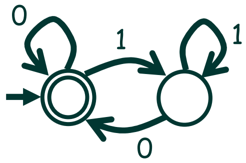
- $\delta^{(k)} : Q\times\Sigma^k \to Q$ mit

  $\delta^{(0)}(q, \varepsilon) := q$

  $\delta^{(i+1)}(q, wx) := \delta(\delta^{(i)}(q, w), x)$ für $w\in\Sigma^i, x\in\Sigma$

  k-Schritt Zustandsüberführungsfunktion und
- $\delta^* : Q\times\Sigma^* \to Q$ mit $\delta^*(q, w) := \delta^{(|w|)}(q, w)$

  iterierte Zustandsüberführungsfunktion



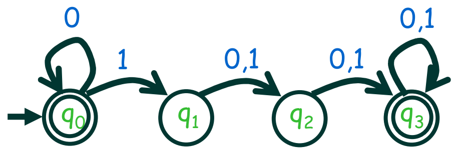
Akzeptanzkriterium

Def.: Sei $A=(Q,\Sigma,\delta,q_0,F)$ dea.

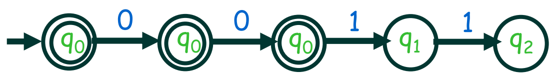
- A akzeptiert $w\in\Sigma^*$, falls $\delta^*(q_0,w)\in F$
- $L(A) := \{w\in\Sigma^* \mid A \text{ akzeptiert } w\}$

  heißt "die von A akzeptierte Sprache"
- $L\subset\Sigma^*$ heißt dea-Sprache, falls es einen dea A gibt mit $L(A)=L$

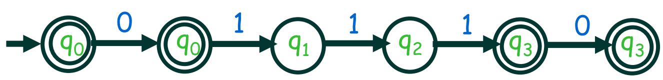
Bsp.: Deterministische endliche Automaten



- Eingabe $w = 0011 \in L(A)^c$



- Eingabe $w = 01110 \in L(A)$

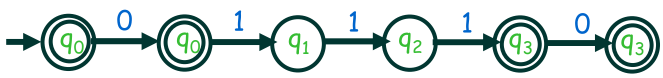


Berechnungspfad

Def.: Sei $A=(Q,\Sigma,\delta,q_0,F)$ dea. Der Berechnungspfad von A auf $w=w_1w_2...w_n \in \Sigma^*$ ist die Folge der Zustände $q_0, q_1,..., q_n$ die A bei der Abarbeitung von w durchläuft.

- gerichteter Pfad $q_{i+1} = \delta(q_i,w_{i+1})$
- Wurzel q_0
- Blatt $\delta^*(q_0,w)$

Bsp.: Eingabe $w = 01110 \in L(A)$



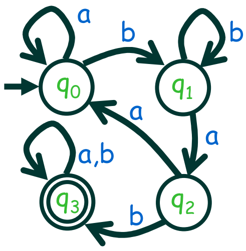
Bsp.: Deterministische endliche Automaten

Bsp.: Ist $L = \{w \in \{a,b\}^* \mid w$ enthält bab als Teilwort$\}$ eine dea-Sprache?

Idee: mit 4 Zuständen können wir zählen, wieviel von bab wir schon gesehen haben (= Länge des Präfixes)

→ Zustände $q_0,\dots,q_3$ wobei

A in q_i, falls Präfix der Länge i gelesen wurde

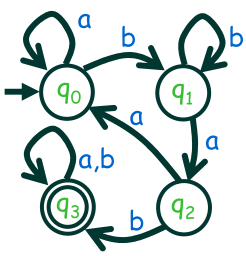


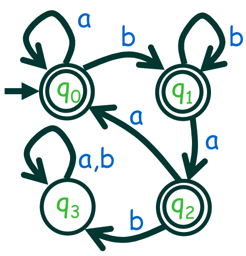
1. abaababaab $\in L(A)$
2. abaaba $\notin L(A)$

Abschlusseigenschaften von dea

- **Satz:** L ist dea-Sprache gdw. $L^c := \Sigma^*\backslash L$ ist dea-Sprache
- **Bew.:** $A=(Q,\Sigma,\delta,q_0,F)$ dea mit $L(A)=L$

  → $A^c:=(Q,\Sigma,\delta,q_0,Q\backslash F)$ dea mit $L(A^c)=L^c$
- **Bsp.:**





Reguläre Sprachen

- **Def.:** Formale Sprache $L \subset \Sigma^*$ heißt regulär, wenn
  - entweder
    - $L = \{\}$,
    - $L = \{\varepsilon\}$, oder
    - $L = \{x\}$ für $x \in \Sigma$

      ("*elementare reguläre Sprachen*")
  - oder $L_1, L_2 \subset \Sigma^*$ sind regulär und
    - $L = L_1 \cup L_2$,
    - $L = L_1 \circ L_2$, oder
    - $L = L_1^*$

  L läßt sich mit endlich vielen regulären Operationen aus den elementaren regulären Sprachen ableiten

Reguläre Sprachen

- **Bsp.:**
  - jede endliche Sprache ist regulär
  - $L = (\{0\} \cup \{1\})^* \circ (\{0\} \circ (\{0\} \cup \{1\}))$

    Wörter über $\{0,1\}$ mit 0 an vorletzter Stelle

Reguläre Ausdrücke - Syntax

- **Def.:** Alphabet Σ' mit $\{\,\underline{(}\,,\underline{)}\,,\underline{\cup}\,,\underline{\circ}\,,\underline{*}\,,\underline{\varepsilon}\,\} \cap \Sigma' = \{\}$
  Ein Wort $a \in \Sigma^*$ wobei $\Sigma := \Sigma' \cup \{\,\underline{(}\,,\underline{)}\,,\underline{\cup}\,,\underline{\circ}\,,\underline{*}\,,\underline{\varepsilon}\,\}$ heißt
  regulärer Ausdruck über Σ', wenn
  - entweder
    - $a = \varepsilon$
    - $a = \underline{\varepsilon}$
    - $a = x$ für $x \in \Sigma'$

      ("*elementare reguläre Ausdrücke*")
  - oder $a_1, a_2 \in \Sigma^*$ sind reguläre Ausdrücke und
    - $a = \underline{(}a_1 \underline{\cup} a_2\underline{)}$
    - $a = \underline{(}a_1 \underline{\circ} a_2\underline{)}$
    - $a = \underline{(}a_1\underline{*}\underline{)}$

Reguläre Ausdrücke - Semantik

- **Def.:** Reguläre Ausdrücke $a, a_1, a_2 \in \Sigma^*$ über Σ'
  Die von a definierte Sprache $L(a) \subset \Sigma'^*$ ist
    - $L(a) := \{\}$ falls $a = \varepsilon$
    - $L(a) := \{\varepsilon\}$ falls $a = \underline{\varepsilon}$
    - $L(a) := \{x\}$ falls $a = x$ für $x \in \Sigma'$
    - $L(a) := L(a_1) \cup L(a_2)$ falls $a = \underline{(}a_1 \underline{\cup} a_2\underline{)}$
    - $L(a) := L(a_1) \circ L(a_2)$ falls $a = \underline{(}a_1 \underline{\circ} a_2\underline{)}$
    - $L(a) := L(a_1)^*$ falls $a = \underline{(}a_1\underline{*}\underline{)}$

- **Satz:** a regulärer Ausdruck → $L(a)$ ist regulär
- **Bew.:** → *Übung*

Reguläre Ausdrücke/Sprachen

- **Bem.**:
  - verwenden x an Stelle von $\underline{x}$
  - statt $\cup$ wird oft + oder | geschrieben
  - $\circ$ wird oft weggelassen
  - Klammern werden oft weggelassen

    Operatorpräzedenz: $* > \circ > \cup$

    $\varepsilon \cup 00^*1^*$ statt $\underline{\varepsilon \cup (0 \circ ((0)^* \circ (1)^*))}$

Beispiele regulärer Ausdrücke/Sprachen

- $a_1 = (0\cup 1)^* \circ 0 \circ (0\cup 1)$

  $L(a_1)$ = Wörter über $\{0,1\}$ mit 0 an vorletzter Stelle

- $a_2 = (0\cup 1)^*10(0\cup 1)^*$

  $L(a_2) = \{w\in\{0,1\}^* \mid w$ enthält 10 als Teilwort$\}$

- $a_3 = 0^*1^*$

  $L(a_3) = \{w\in\{0,1\}^* \mid w$ enthält 10 nicht als Teilwort$\} = L(a_2)^c$

- $a_4 = (0\cup 1)^*101(0\cup 1)^*$

  $L(a_4) = \{w\in\{0,1\}^* \mid w$ enthält 101 als Teilwort$\}$

Beispiele regulärer Ausdrücke/Sprachen

- **Frage**: Gibt es regulären Ausdruck a_5 mit
  $L(a_5) = \{w \in \{0,1\}^* \mid w$ enthält 101 nicht als Teilwort$\} = L(a_4)^c$
  Fall 1.: w enthält nichtmal 10 als Teilwort
  Fall 2.: w enthält 10 n-mal als Teilwort
  → $w = w_1 10 w_2 10 \ldots w_n 10 w_{n+1}$ wobei
  $w_2,\ldots,w_{n+1}$ beginnen nicht mit 1,
  w_1,w_{n+1} können leer sein
  $w_2,\ldots,w_n$ sind nicht leer
  $w_1,\ldots,w_n$ enthalten kein 10 also
  $w = v_1 \circ 100 \circ v_2 \circ 100 \circ \ldots \circ v_n \circ 10 \circ v_{n+1}$ mit
  $v_1,\ldots,v_n \in 0^*1^*$ und $v_{n+1} \in \varepsilon \cup 0 \circ 0^*1^*$
  also $a_5 = 0^*1^* \cup ((0^*1^*100)^*0^*1^*10(\varepsilon \cup 0 \circ 0^*1^*))$

Komplement regulärer Sprachen

- **Frage**: Gibt es zu jedem regulären Ausdruck a einen regulären Ausdruck a^c mit $L(a^c) = L(a)^c$

  "Sind reguläre Sprachen unter Komplementbildung abgeschlossen?"

- **Antwort**: Ja, weil dea-Sprachen = reguläre Sprachen

- **Zunächst**: Reguläre Sprachen sind dea-Sprachen

Reguläre Sprachen sind dea-Sprachen

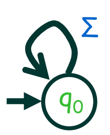
- **Idee**: induktiver "Nachbau" eines regulären Ausdrucks a als endlicher Automat $M(a)$, so dass $L(a) = L(M(a))$

  Sprache $L(a)c\Sigma^*$ sei beschrieben durch einen regulären Ausdruck a mit n Operationen aus $\{\cup, \circ, *\}$

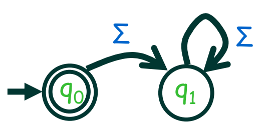
Reguläre Sprachen sind dea-Sprachen

- n=0

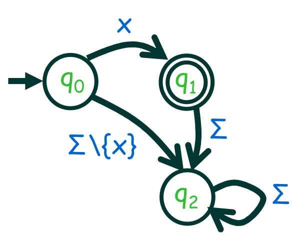
$a = \varepsilon$
$L(a) = \{\}$



$a = \underline{\varepsilon}$
$L(a) = \{\varepsilon\}$



$a = x$ für $x \in \Sigma$
$L(a) = \{x\}$



Reguläre Sprachen sind dea-Sprachen

- <u>$n \rightarrow n+1$</u>

  Sprachen $L(a_1), L(a_2) \subset \Sigma^*$ seien beschrieben durch je einen regulären Ausdruck a_1, a_2 mit $\leq n$ Operationen aus $\{\cup, \circ, *\}$

  → Nach Induktionsvoraussetzung gibt es dea M_1, M_2 mit $L(a_1) = L(M_1)$ und $L(a_2) = L(M_2)$ d.h. $L(a_1)$ und $L(a_2)$ sind dea-Sprachen

  zu zeigen:

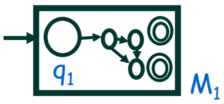
  - $L(a_1) \cup L(a_2)$
  - $L(a_1) \circ L(a_2)$
  - $L(a_1)^*$

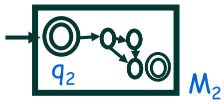
  sind dea-Sprachen

Reguläre Sprachen sind dea-Sprachen

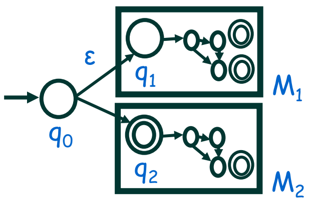
<u>$L(a_1) \cup L(a_2)$</u>

$M_1=(Q_1,\Sigma,\delta_1,q_1,F_1)$ $M_2=(Q_2,\Sigma,\delta_2,q_2,F_2)$





Idee: neuer Startzustand q_0 der bei der Verarbeitung von $w \in \Sigma^*$ beide Startzustände q_1 und q_2 aktiviert, ohne dabei das erste Zeichen von w zu lesen



Problem:

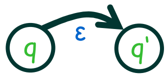
sprengt den Rahmen des dea-Modelles!

Reguläre Sprachen sind nea-Sprachen

Idee: Erweiterung des dea-Modells, so dass

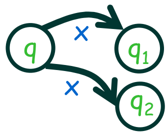
- von einem Zustand q der Übergang in einen Nachfolgezustand q' möglich ist, ohne ein Zeichen der Eingabe zu lesen

  ("ε -Übergänge")



- von einem Zustand q beim Lesen eines Symbols x der Übergang in einen von mehreren möglichen Nachfolgezuständen $q_1,\dots,q_k$ möglich ist

  (" Nichtdeterminismus")



"nichtdeterministische endliche Automaten mit ε -Übergängen"