

Einführung in die Computerlinguistik

Pumping-Lemma für reguläre
Sprachen

Abschlußeigenschaften regulärer Sprachen

Vereinigung	+ ✓
Schnittmenge	+
Komplementmenge	+
Konkatenation	+ ✓
Kleene Stern	+ ✓
Schnittmenge mit einer regulären Sprache	+

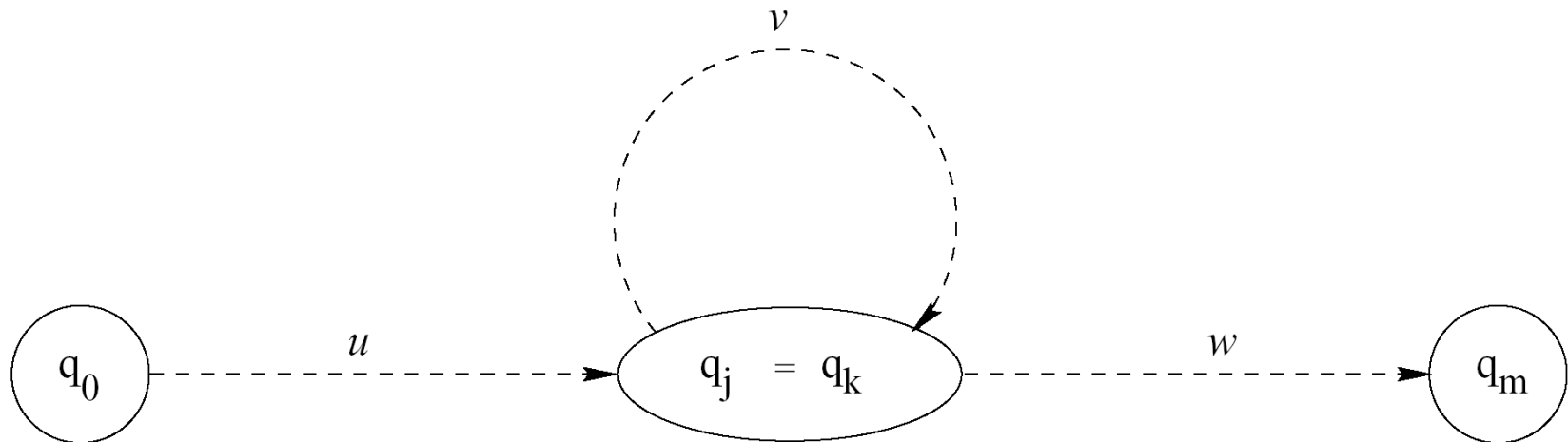
Komplement: konstruiere komplementären DEA

Schnittmenge: folgt aus *de Morgan*

Pumping-Lemma für reguläre Sprachen

Lemma 9. [Pumping-Lemma] Wenn L eine unendliche reguläre Sprache ist, dann gibt es Wörter $u, v, w \in \Sigma^*$, so daß $v \neq \epsilon$ und $uw^i w \in L$ für beliebiges $i \geq 0$.

Beweisskizze:



$L = \{a^n b^n : n \geq 0\}$ ist nicht regulär

$L = \{a^n b^n : n \geq 0\}$:

L ist unendlich. Wäre L regulär, dann gäbe es $u, v, w \in \{a, b\}^*$, $v \neq \epsilon$ mit $uv^n w \in L$ für alle $n \geq 0$. Für v ergeben sich folgende Fälle:

- angenommen v besteht aus a 's und b 's, dann gibt es in v^n ($n \geq 2$) b 's, die vor a 's stehen (Widerspruch),
- angenommen v besteht nur aus a 's, dann sind alle b 's in dem w -Teil; wenn man v aufpumpt, steigt die Zahl der a 's ohne daß die der b 's steigen würde (Widerspruch).
- angenommen v besteht nur aus b 's, dann analog zu v besteht nur aus a 's,

Natürliche Sprachen sind nicht regulär (1)

1. Der Hund starb.
 2. Der Hund, der den Vogel jagte, starb.
 3. Der Hund, der den Vogel, der den Wurm fraß, jagte, starb.
 4. Der Hund, der den Vogel, der den Wurm, der den Rasen durchquerte, fraß, jagte, starb.
 5. Der Hund, der den Vogel, der den Wurm, der den Rasen, der den Garten bedeckte, durchquerte, fraß, jagte, starb.
- ...

- Allgemeine Form: der Hund (der den *maskulines Nomen*)ⁿ (*transitives Verb*)ⁿ starb.

Natürliche Sprachen sind nicht regulär (2)

- Sei $A = \left\{ \begin{array}{l} \text{der den Hund, der den Vogel, der den Kühlschrank,} \\ \text{der den Wurm, der den Rasen, der den Garten} \end{array} \right\}$ und
- $B = \{\text{fraß, beschenkte, durchquerte, jagte, liebte, sah, trank}\}$ und
- $w = \text{der Hund}$ und $v = \text{starb}$.
- wx^*y^*v mit $x \in A$ und $y \in B$ ist eine reguläre Sprache.
- $\text{DEUTSCH} \cap wx^*y^*v = wx^n y^n v$.
- Wäre DEUTSCH regulär, dann müßte auch $wx^n y^n v$ regulär sein, da die Schnittmenge zweier regulärer Sprachen regulär ist (Widerspruch).

intuitive Merkgeregeln über reguläre Sprachen

- L ist regulär, wenn der Test auf Mitgliedschaft eines Wortes w in L durch buchstabenweises Durchlesen unter Zuhilfenahme eines beschränkten Speichers möglich ist.
- Endliche Automaten sind zu schwach für:
 - Worteigenschaften, deren Überprüfung eine Zählervariable mit Wertebereich \mathbb{N} erfordert
 - Spezifikation der Wiederholung fester Muster beliebiger Länge
 - Mengen von Ausdrücken mit beliebiger Klammertiefe

Übungsaufgaben

Welche Aussagen kann man mit Hilfe der Abschlußeigenschaften der regulären Sprachen und dem Pumping-Lemma über die Komplexität folgender formaler Sprachen machen:

1. $L_1 = \{w \in \{a, b\}^* : w \text{ enthält eine ungerade Anzahl von } b's\}$.
2. $L_2 = \{w \in \{a, b\}^* : w \text{ enthält die gleiche Anzahl von } b's \text{ und } a's\}$.
3. w^R ist das Wort w in umgekehrter Reihenfolge.
 $L_3 = \{ww^R : w \in \{a, b\}^*\}$.