Reguljära uttryck

- Vi har tittat på två beräkningsmodeller som båda tillåter oss att representera de reguljära språken, nämligen deterministiska och ickedeterministiska finita automater.
- En tredje representationstyp är *Reguljära uttryck*.
- Innebär att man bygger upp språk med hjälp av ett alfabet Σ, konstanterna ϵ och $\emptyset$, samt de reguljära operationerna union, konkatenation och stjärna.
- Jämför t.ex. termen $((5+3) \times 4)$ med $((0 \cup 1) \circ (0^*))$.
- Reguljära uttryck har ni troligtvis redan stött på, t.ex. i Awk, Grep, Vim, eller Perl.

Reguljära uttryck - syntax

Definition 1.52 Vi säger att R är ett reguljärt uttryck om R är ...

- a, för något a i alfabetet Σ,
- ϵ,
- $\emptyset$,
- $(R_1 \cup R_2)$, där R_1 och R_2 är reguljära uttryck,
- $(R_1 \circ R_2)$, där R_1 och R_2 är reguljära uttryck, eller
- (R_1^*), där R_1 är ett reguljärt uttryck.

Exempel: Vi låter Σ vara mängden $\{a, b, c\}$, då är $((b \cup c) \circ \emptyset)$, $(((\epsilon \circ a) \circ \epsilon)^*)$ och $((a^*) \cup (b \circ c))$ reguljära uttryck.

Förenklad notation

- Vi använder prioritetsordningen stjärna, konkatenation, och till sist union vid evaluering.
- Då $\cup$ och $\circ$ är associativa kan vi utelämna paranteserna utan risk för förvirring.
- Dessutom skriver vi vanligtvis $R_1 R_2$ istället för $R_1 \circ R_2$,
- samt R^+ för RR^*,
- och R^k för konkatenationen av k stycken R med varandra.
- Det innebär att uttrycket $(((a \circ a) \circ a) \cup (b \circ (b^*)))$ får skrivas $a^3 \cup b^+$, vilket är mer lättläst.

Semmantik

Varje reguljärt uttryck E definierar ett språk $L(E)$:

- $L(a) = \{a\}$
- $L(\epsilon) = \{\epsilon\}$
- $L(\emptyset) = \emptyset$
- $L(E \cup E') = L(E) \cup L(E')$
- $L(E \circ E') = L(E) \circ L(E')$
- $L(E^*) = L(E)^*$

Exempel

Vi skriver språket som representeras av det reguljära uttrycket R som $L(R)$. Om Σ är $\{1, 0\}$, vad är då $L(R)$ om R är ...

$(0 \cup 1)^k$	$0^*10^*10^*$
$(\Sigma\Sigma)^*\Sigma$	$10\emptyset$
10ϵ	$0\Sigma^*1 \cup 1\Sigma^*0 \cup \epsilon$
$0\Sigma^*1 \cup 1\Sigma^*0 \cup \emptyset$	$(1 \cup 11 \cup 111)(0 \cup 00 \cup 000)\epsilon$
$1^*\emptyset$	$\emptyset^*$

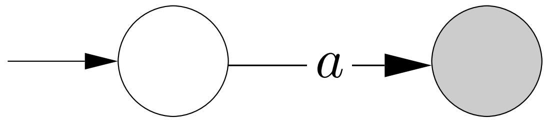
Ekvivalens med finit automata

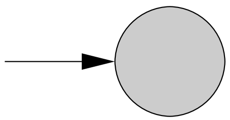
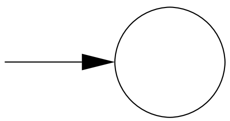
Teorem 1.54 Ett språk är reguljärt om och endast om det finns ett reguljärt uttryck som beskriver det.

Ekvivalens med finit automata

Lemma 1.55 Ett språk är reguljärt om det finns ett reguljärt uttryck som beskriver det.

Bevis Enligt definitionen kan ett reguljärt uttryck R ha sex olika former, och för var och en av dess går det att konstruera en språkekvivalent NFA.

- $R = a$ för något a i Σ motsvaras av  ,

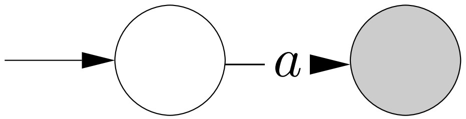
- $R = \epsilon$ av  , och
- $R = \emptyset$ av  .

Automaterna för $R = R_1 \cup R_2$, $R = R_1 R_2$, och $R = R_1^*$ konstrueras som vanligt.

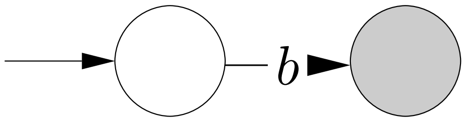
Exempel

Hitta en finit automat som accepterar $L((a \cup b)^* a\emptyset$.

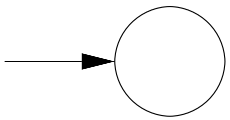
a



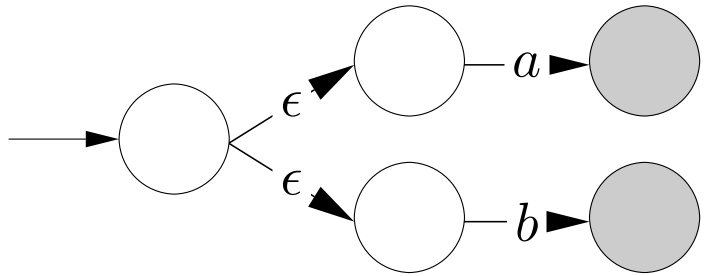
b



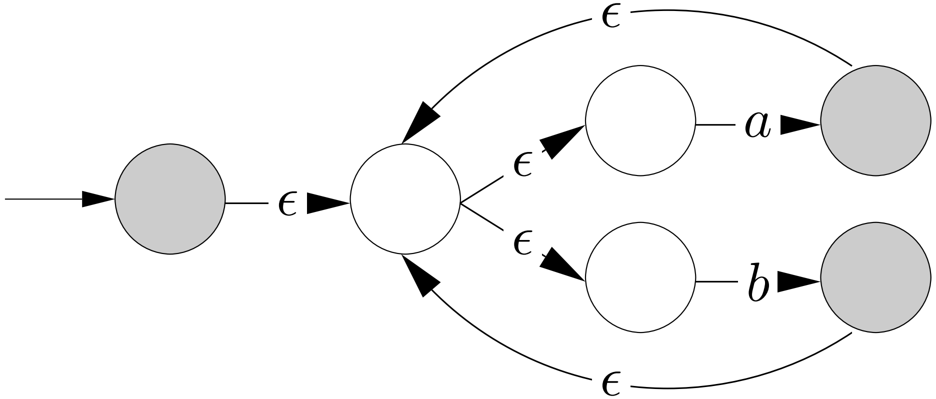
$\emptyset$



$a \cup b$



$(a \cup b)^*$



Exempel

Den resulterande automaten blir som nedan. Hur många strängar finns i det accepterade språket?