

CS_275 Automata and Formal Language Theory

Course Notes

Part II: The Recognition Problem (II)

Additional Material

(This material is no longer taught and not exam relevant)

Sect II.2.: Basics of Regular Languages and Expressions

Anton Setzer

(Based on a book draft by J. V. Tucker and K. Stephenson)

Dept. of Computer Science, Swansea University

[http://www.cs.swan.ac.uk/~csetzer/lectures/
automataFormalLanguage/current/index.html](http://www.cs.swan.ac.uk/~csetzer/lectures/automataFormalLanguage/current/index.html)

March 17, 2017

II.2.1. Regular Languages (12.2)

II.2.2. Regular Expressions (13.8)

II.2.1. Regular Languages (12.2)

II.2.2. Regular Expressions (13.8)

Proof of Lemma II.2.1.2.

Lemma (II.2.1.2.)

1. Assume a grammar G which has only productions of the form

$$A \longrightarrow Bw \text{ or } A \longrightarrow w$$

for some $w \in T^*$, $A, B \in N$. Then $L(G) = L(G')$ for some left-linear grammar G' , which can be computed from G .

2. Assume a grammar G which has only productions of the form

$$A \longrightarrow wB \text{ or } A \longrightarrow w$$

for some $w \in T^*$, $A, B \in N$. Then $L(G) = L(G')$ for some right-linear grammar G' , which can be computed from G .

Proof of Lemma II.2.1.2.

In a first step we omit all transitions $A \rightarrow B$ for $A, B \in N$:

Let $G = (N, T, S, P)$ be a grammar having such transitions.

We form a grammar G' having no such transitions as follows:

grammar	G'
terminals	N
nonterminals	T
start symbol	S
productions	$A \rightarrow w$ if $A \Rightarrow_G^* A' \rightarrow w$ for some $A, A' \in N, w \in T^*$ $A \rightarrow wB$ if $A \Rightarrow_G^* A' \rightarrow wB$ for some $A, A', B \in N,$ $w \in T^*$

Proof

So in G' we just jump over all silent transitions $A \rightarrow B$ in G .

We can in fact decide whether $A \Rightarrow^* A'$, since such a derivation must have the form $A = A'$ or $A = A_1 \Rightarrow A_2 \Rightarrow \cdots \Rightarrow A_n = A'$ for some $A_i \in N$.

And if such derivation exists then a derivation exists in which all A_i are distinct (omit loops).

Therefore n can be restricted to the number of elements in N , and therefore there are only finitely many possible derivations, which we can enumerate.

For each of them we can check whether it is in fact a derivation, and therefore determine all possible derivations $A \Rightarrow^* A'$.

Proof

Now one can easily see that for $w \in T^*$

$$S \Rightarrow_G^* w \text{ iff } S \Rightarrow_{G'}^* w$$

End of Proof of II.2.1.2.

We have now obtained a grammar which doesn't contain silent productions of the form $A \rightarrow B$ for nonterminals A, B .

The following lemma shows that such languages are definable by left-linear or right-linear grammars.

Proof of Lemma II.2.1.3.

Lemma (II.2.1.3.)

1. Assume a grammar G which has only productions of the form

$$A \longrightarrow Bw \text{ or } A \longrightarrow w'$$

for some $w \in T^+$, $w' \in T^*$, $A, B \in N$. Then $L(G) = L(G')$ for some left-linear grammar G' , and G' can effectively computed from G .

2. Assume a grammar G which has only productions of the form

$$A \longrightarrow wB \text{ or } A \longrightarrow w'$$

for some $w \in T^+$, $w' \in T^*$, $A, B \in N$. Then $L(G) = L(G')$ for some right-linear grammar G' , and G' can effectively computed from G .

Proof of Lemma II.2.1.3.

- ▶ In (2) replace
 - ▶ Productions $A \rightarrow a_1 a_2 \cdots a_n B$ with $n \geq 2$ by $A \rightarrow a_1 A_1$,
 $A_1 \rightarrow a_2 A_2, \dots, A_{n-1} \rightarrow a_n B$ for some new nonterminals A_i .
 - ▶ Productions $A \rightarrow a_1 a_2 \cdots a_n$ with $n \geq 2$ by $A \rightarrow a_1 A_1, A_1 \rightarrow a_2 A_2$
 $, \dots, A_{n-1} \rightarrow a_n$ for some new nonterminals A_i .
- ▶ (1) is proved similarly.

Derivations in Regular Grammars

Theorem

- (a) Let $G = (N, T, S, P)$ be a left-linear grammar, $A \in N$, $w \in (N \cup T)^*$, $A \Rightarrow^* w$.

Then the derivation of $A \Rightarrow^* w$ is

$$A \Rightarrow A_1 a_1 \Rightarrow A_2 a_2 a_1 \Rightarrow \cdots \Rightarrow A_n a_n \cdots a_2 a_1 = w \quad (1)$$

$$\text{or } A \Rightarrow A_1 a_1 \Rightarrow A_2 a_2 a_1 \Rightarrow \cdots \Rightarrow A_n a_n \cdots a_2 a_1 \quad (2)$$

$$\Rightarrow a_{n+1} a_n \cdots a_2 a_1 = w$$

$$\text{or } A \Rightarrow A_1 a_1 \Rightarrow A_2 a_2 a_1 \Rightarrow \cdots \Rightarrow A_n a_n \cdots a_2 a_1 \quad (3)$$

$$\Rightarrow a_n \cdots a_2 a_1 = w$$

for productions

▶ $A_i \rightarrow A_{i+1} a_{i+1}$ (in (1) – (3)),

▶ $A_n \rightarrow a_{n+1}$ (in (2))

▶ $A_n \rightarrow \epsilon$ (in (3))

Derivations in Regular Grammars

Theorem

(b) Let $G = (N, T, S, P)$ be a right-linear grammar, $A \in N$, $w \in (N \cup T)^*$, $A \Rightarrow^* w$.

Then the derivation of $A \Rightarrow^* w$ is

$$A \Rightarrow a_1 A_1 \Rightarrow a_1 a_2 A_2 \Rightarrow \cdots \Rightarrow a_1 a_2 \cdots a_n A_n = w \quad (1)$$

$$\text{or } A \Rightarrow a_1 A_1 \Rightarrow a_1 a_2 A_2 \Rightarrow \cdots \Rightarrow a_1 a_2 \cdots a_n A_n \quad (2)$$

$$\Rightarrow a_1 a_2 \cdots a_n a_{n+1} = w$$

$$\text{or } A \Rightarrow a_1 A_1 \Rightarrow a_1 a_2 A_2 \Rightarrow \cdots \Rightarrow a_1 a_2 \cdots a_n A_n \quad (3)$$

$$\Rightarrow a_1 a_2 \cdots a_n = w$$

for productions

$$\blacktriangleright A_i \longrightarrow a_{i+1} A_{i+1} \text{ (in (1) - (3))}$$

$$\blacktriangleright A_n \longrightarrow a_{n+1} \text{ (in (2))}$$

$$\blacktriangleright A_n \rightarrow \epsilon \text{ (in (3)).}$$

Proof

The above are the only derivations possible.

II.2.1. Regular Languages (12.2)

II.2.2. Regular Expressions (13.8)

Proof of Lemma II.2.2.1.

Lemma (II.2.2.1.)

Let G, G' be both left-linear grammars or both right-linear grammars. Then we can define a left-linear or right-linear grammars G_i s.t.

1. $L(G_1) = L(G) \mid L(G')$,
2. $L(G_2) = L(G).L(G')$,
3. $L(G_3) = L(G)^*$.

These grammars can be computed from G and G' .

Proof of Lemma II.2.2.1.

Assume in 1./2./3.

$$G = (T, N, S, P) , \quad G' = (T', N', S', P') .$$

After renaming of nonterminals we can assume $N \cap N' = \emptyset$.

Let S'' be a new symbol not in $N \cup N' \cup T \cup T'$.

We define multi-step left/right-linear grammars with those properties, from which one can construct ordinary (one-step) left/right-linear grammars with those properties.

We only carry out the proof for right-linear grammars.

Proof of 1.

We define G_1 as follows:

grammar	G_1
terminals	$T \cup T'$
nonterminals	$N \cup N' \cup \{S''\}$
start symbol	S''
productions	$S'' \rightarrow S$ $S'' \rightarrow S'$ P P'

Proof of 1.

So G_1 has the productions from G and G' plus

$$S'' \rightarrow S \text{ and } S'' \rightarrow S' .$$

Derivations in G_1 have the form

$$S'' \Rightarrow S \Rightarrow^* w$$

and

$$S'' \Rightarrow S' \Rightarrow^* w'$$

for derivations

$$S \Rightarrow_G^* w$$

and

$$S' \Rightarrow_{G'}^* w'$$

So for $w'' \in (T \cup T')^*$ we have

$$S'' \Rightarrow_{G_1}^* w'' \text{ iff } S \Rightarrow_G^* w'' \text{ or } S' \Rightarrow_{G'}^* w'',$$

so $L(G_1) = L(G) \cup L(G')$.

Proof of 2.

We define G_2 as follows:

grammar	G_2
terminals	$T \cup T'$
nonterminals	$N \cup N'$
start symbol	S
productions	$A \rightarrow aA'$ for $A \rightarrow aA' \in P$ ($A, A' \in N, a \in T$) $A \rightarrow aS'$ for $A \rightarrow a \in P$ ($A \in N, a \in T$) P'

Proof of 2.

So G_2 has

- ▶ the productions from G' ,
- ▶ the productions of the form $A \rightarrow aA$ from G and
- ▶ productions $A \rightarrow aS'$, if $A \rightarrow a$ is a production from G .

A derivation in G_2 starts with a derivation

$$S \Rightarrow a_1 A_1 \Rightarrow a_1 a_2 A_2 \Rightarrow a_1 a_2 a_3 A_3 \Rightarrow \cdots \Rightarrow a_1 a_2 \cdots a_{n-1} A_{n-1} \\ \Rightarrow a_1 a_2 \cdots a_n S'$$

for derivations in G of the form

$$S \Rightarrow a_1 A_1 \Rightarrow a_1 a_2 A_2 \Rightarrow a_1 a_2 a_3 A_3 \Rightarrow \cdots \Rightarrow a_1 a_2 \cdots a_{n-1} A_{n-1} \\ \Rightarrow a_1 a_2 \cdots a_n .$$

Proof of 2.

Then this is followed by a derivation

$$\begin{aligned} a_1 a_2 \cdots a_n S' &\Rightarrow a_1 a_2 \cdots a_n b_1 B_1 \Rightarrow a_1 a_2 \cdots a_n b_1 b_2 B_2 \Rightarrow \cdots \\ &\Rightarrow a_1 a_2 \cdots a_n b_1 b_2 \cdots b_{m-1} B_{m-1} \Rightarrow a_1 a_2 \cdots a_n b_1 b_2 \cdots b_m, \end{aligned}$$

for a derivation in G' of the form

$$\begin{aligned} S' &\Rightarrow b_1 B_1 \Rightarrow b_1 b_2 B_2 \Rightarrow \cdots \\ &\Rightarrow b_1 b_2 \cdots b_{m-1} B_{m-1} \Rightarrow b_1 b_2 \cdots b_m \end{aligned}$$

Therefore $S \Rightarrow_{G_2}^* w$ for some $w \in (T \cup T')^*$ if and only if $S \Rightarrow_{G_1}^* w'$ and $S' \Rightarrow_{G_2}^* w''$ for some w', w'' s.t. $w = ww''$.

So $L(G_2) = L(G).L(G')$.

Proof of 3.

We define G_3 as follows:

grammar	G_3
terminals	T
nonterminals	N
start symbol	S
productions	$S \rightarrow \epsilon,$ $A \rightarrow aA'$ for $A \rightarrow aA' \in P$ ($A, A' \in N, a \in T$) $A \rightarrow aS$ for $A \rightarrow a \in P$ ($A \in N, a \in T$)

Proof of 3.

Derivations in G_3 are $S \Rightarrow \epsilon$ or they start similarly as for concatenation with

$$S \Rightarrow^* wS$$

for a derivation in G

$$S \Rightarrow^* w$$

and $w \in N^+$. In the latter case it can continue either (using $S \rightarrow \epsilon$) with $wS \Rightarrow w$ or with

$$wS \Rightarrow^* ww'S$$

for a derivation in G

$$S \Rightarrow^* w'$$

Again in the latter case we can continue (using $S \rightarrow \epsilon$) with $ww'S \rightarrow ww'$ or with

$$ww'S \Rightarrow^* ww'w''S$$

for a derivation in G

$$S \Rightarrow^* w''$$

Proof of 3.

We obtain that in G_3 we have

$$S \Rightarrow^* w$$

if there exist derivations in G of

- ▶ $S \Rightarrow^* w_1$
- ▶ $S \Rightarrow^* w_2$
- ▶ ...
- ▶ $S \Rightarrow^* w_n$

s.t. $w = w_1 w_2 \cdots w_n$. So we get

$$L(G_3) = \{w_1 w_2 \cdots w_n \mid n \geq 0, w_1, \dots, w_n \in L(G)\} = L(G)^*$$

Proof of Lemma II.2.2.2.

Lemma (II.2.2.2.)

Let E be a regular Expression. Then there exist both left-linear and right-linear grammars G, G' s.t.

$$L(E) = L(G) = L(G')$$

G and G' can be computed from L .

Proof: By Lemma II.2.2.1, and the fact that the finite languages $\emptyset, \{\epsilon\}$ and $\{a\}$ are regular.

Proof of Lemma II.2.2.2.

Induction on the definition of regular expressions.

Case 1: $L = \emptyset, \epsilon, \mathbf{a}$

(where $a \in T$). Then L is finite, therefore definable by a left/right-linear grammar.

Case 2: $L = (L_1) \mid (L_2)$ or $L = (L_1)(L_2)$ or $L = (L_1)^*$. By IH L_i are defined by left/right-linear grammars G_i . By Lemma II.2.2.1. it follows that L can be defined by a left/right-linear grammar.