cse581
Computer Science Fundamentals: Theory

Professor Anita Wasilewska

TCB - LECTURE 5

CONTEXT -FREE GRAMMARS and LANGUAGES

Context-free Grammars

Finite Automata are formal language **recognizers**

- they are devises that **accept** valid strings

Context-free Grammars are a certain type of formal **language generators**

- they are devises that **produce** valid strings

Context-free Grammars

Such a language generator **devise** begins,
when given a start symbol, to **construct** a string

Its operation is not completely determined from he beginning
but is nevertheless limited by a **finite** set of rules

The process **stops**,
and the **devise** outputs a completed string

The language defined by the **devise** is the set of all strings
it can **produce**

Context-free Grammar Definition

Definition

A Context-Free Grammar is a quadruple

$$G = (V, \Sigma, R, S)$$

where

V is an **alphabet**

$\Sigma \subseteq V$ is a set of **terminals**

$V - \Sigma$ is the set of **nonterminals**

R is a finite set of **rules**

$$R \subseteq (V - \Sigma) \times V^*$$

$S \in V - \Sigma$ is the **start symbol**

Context-free Grammar Definitions

The alphabet V consists of two disjoint parts: **nonterminals** $V - \Sigma$ and **terminals** Σ, i.e.

$$V = (V - \Sigma) \cup \Sigma$$

Notations

We use symbols of capital letters, with indices if necessary for **nonterminals** $V - \Sigma$, i.e.

$$A,\ B, C, S, T, X, Y, \dots A_i,\ \dots \in V - \Sigma$$

The **terminal** alphabet Σ is as in case of the finite automata, the alphabet the words of the language are made from and we denote its elenets, as before by small letters, or symbol σ, with indices if necessary, i.e.

$$a,\ b, c,\ \sigma, \dots a_i, \dots \sigma_i,\ \dots \in \Sigma$$

Context-free Grammar Definition

Notations

By definition, the set of **rules** R of a context-free grammar G is a **finite** set such that

$$R \subseteq (V - \Sigma) \times V^*$$

It means that $R = \{(A, u) : \quad A \in (V - \Sigma) \text{ and } u \in V^*\}$

where A is a **nonterminal** and $u \in V^*$ is a string that contains some **terminals** and **nonterminals**

We write

$$A \to_G u \quad \text{or} \quad A \to u \quad \text{for any} \quad (A, u) \in R$$

Context-free Grammar Definition

Given a **context-free grammar**

$$G = (V,\ \Sigma,\ R,\ S)$$

Definition

For any $u, v \in V^*$, we define a **one step derivation**

$$u \underset{G}{\Rightarrow} v$$

of v from u as follows

$\Rightarrow_G$ if and only if there are $A, x, y, v' \in V^*$ such that

1. $A \in V - \Sigma$

2. $u = xAy$ and $v = xv'y$

2. $A \rightarrow v'$ for certain $r \in R$

Context-free Grammar Definition

One step derivation in plain words:

$u \Rightarrow_G v$ if and only if we obtain v from u by a **direct** application of one rule $r \in R$

Definition of the language L(G) **generated** by G

$$L(G) = \{w \in \Sigma^* : S \overset{*}{\underset{G}{\Rightarrow}} w\}$$

where $\Rightarrow_G^*$ is a transitive, reflexive closure of $\Rightarrow_G$

Context-free and Regular Languages

Given a **derivation** of $w \in \Sigma^*$ in G

$$S \underset{G}{\overset{*}{\Rightarrow}} w$$

We write is in detail (by definition of $\Rightarrow^*_G$) as

$$S \underset{G}{\Rightarrow} w_1 \underset{G}{\Rightarrow} w_2 \underset{G}{\Rightarrow} \ldots \underset{G}{\Rightarrow} w \quad \text{for} \quad w_i \in V^*,\; w \in \Sigma$$

or when G is known as

$$S \Rightarrow w_1 \Rightarrow w_2 \Rightarrow \ldots \Rightarrow w$$

or just as a sequence of words

$$S, \; w_1, \; w_2, \; \ldots \; w \quad \text{for} \quad w_i \in V^*,\; w \in \Sigma^*$$

Context-free Grammar Example

Example

Consider a grammar $G = (V,\ \Sigma,\ R,\ S)$

for $V = \{S, a, b\}$, $\Sigma = \{a, b\}$ and

$R = \{r1:\ S \rightarrow aSb,\quad r2:\ S \rightarrow e\}$

Here are some **derivations** in G

D1 $S \Rightarrow e$ so we have that

$$e \in L(G)$$

D2 $S \Rightarrow^{r1} aSb \Rightarrow^{r1} aaSbb \Rightarrow^{r2} aabb$

or we just write the derivation as

$$S,\ aSb,\ aaSbb,\ aabb$$

and we have that

$$aabb \in L(G)$$

Context-free Languages

D3 $S \Rightarrow^{r1} aSb \Rightarrow^{r1} aaSbb \Rightarrow^{r1} aaaSbbb \Rightarrow^{r2} aaabbb$

or we also write the derivation as

$$S, \quad aSb, \quad aaSbb, \quad aaaSbbb, \quad aaabbb$$

and we have that

$$a^3b^3 \in L(G)$$

We prove, by induction on the length of derivation that

$$L(G) = \{a^n b^n : \quad n \geq 0\}$$

Context-free and Regular Languages

Definition

A language L is a **context-free language** if and only if there is a context-free grammar G such that

$$L = L(G)$$

We have just proved

Fact 1

The language $L = \{a^n b^n : \ n \geq 0\}$ is **context-free**

Context-free and Regular Languages

Observe that we also proved that the language

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

is **not regular**

Denote by RL the class of all regular languages and by CFL the class of all contex-free languages

Context-free and Regular Languages

Hence we have **proved**

Fact 2 $RL \neq CFL$

Our next GOAL will be **to prove** the following

Theorem

The the class of all regular languages is a proper subset of the class of all contex-free languages, i.e.

$$RL \subset CFL$$

Exercises

Exercise 1

Show that the regular language $L = \{a^* : \ a \in \Sigma\}$ is **context-free**

Proof By definition of **context-free** language we have to construct a CF grammar G such that

$$L = L(G) \ \text{ i.e } \ L(G) = \{a^* : \ a \in \Sigma\}$$

Here is the grammar $G = (V, \ \Sigma, \ R, \ S)$

for $V = \{S, a\}$, $\Sigma = \{a\}$ and

$R = \{ \ S \ \rightarrow \ aS, \quad S \rightarrow e\}$

We write rules of R in a shorter way as

$$R = \{ \ S \ \rightarrow \ aS \mid e\}$$

Exercises

Here is a formal **derivation** in G:

$$S \Rightarrow aS \Rightarrow aaS \Rightarrow aaaS \Rightarrow aaa$$

or written as a sequence of words

$$S, \quad aS, \quad aaS, \quad aaaS, \quad aa$$

and we have

$$aaaa \in L(G)$$

We prove, by induction on the length of derivation that

$$L(G) = \{a^* : \quad a \in \Sigma\}$$

Exercises

Exercise 2

Show that the NOT regular language

$$L = \{ww^R : \quad w \in \{a, b\}^*\}$$

is **context-free**

Exercises

We construct a context-free grammar G such that

$L(G) = \{ww^R : \quad w \in \{a, b\}^*\}$

as $G = (V, \Sigma, R, S)$

where $V = \{a, b, S\}, \ \Sigma = \{a, b\}$

$$R = \{S \rightarrow aSa \mid bSb \mid e\}$$

Derivation example:

$S \Rightarrow aSa \Rightarrow abSba \Rightarrow abbSbba \Rightarrow abbbba$

or written as $S, \ aSa, \ abSba, \ abbSbba, \ abbbba$

We prove, by induction on the length of derivation that

$$ww^R \in L(G) \text{ for any } w \in \Sigma^*$$

Exercises

Remark

The set of rules

$$R = \{S \;\rightarrow\; aSa \mid aSb \mid c\}$$

defines a grammar G with the language

$$L(G) = \{wcw^R : \quad w \in \{a, b\}^*\}$$

Exercise 3

Show that the NOT regular language

$$L = \{w \in \{a, b\}^* : \quad w = w^R\}$$

is **context-free**

Exercises

We construct a context-free grammar G such that

$$L(G) = \{w \in \{a,b\}^* : \; w = w^R\}$$

as follows

$G = (V, \Sigma, R, S)$, where $V = \{a, b, S\}$, $\Sigma = \{a, b\}$

$$R = \{S \to aSa \,|bSb \mid a \mid b \mid e\}$$

Derivation example: $S \Rightarrow aSa \Rightarrow abSba \Rightarrow ababa$

We check

$$(ababa)^R = ((aba)(ba))^R = (ba)^R((ab)a)^R = aba(ab)^R = ababa$$

We used **Property**: for any $x, y \in \Sigma^*$, $(xy)^R = y^R x^R$ and

Definition: for any $x \in \Sigma^*$, $a \in \Sigma$, $e^R = e$, $(xa)^R = ax^R$

Exercises

Grammar correctness justification

Observe that the rules

$$S \to aSa \mid bSb \mid e$$

generate the language (as was proved in Example 2)

$$L_1 = \{ww^R : \; w \in \Sigma^*\}$$

Adding additional rules $S \to a \mid b$ we get that

$$L(G) = L_1 \cup \{waw^R : \; w \in \Sigma^*\} \cup \{wbw^R : \; w \in \Sigma^*\}$$

Hence

$w \in L(G)$ iff $w = xx^R$ or $w = xax^R$ or $w = xbx^R$

Exercises

Hence

$w \in L(G)$ iff $w = xx^R$ or $w = xax^R$ or $w = xbx^R$

We show now that in each case $w = w^R$, i.e.

we **prove** that

$$L(G) = \{w \in \{a, b\}^* : \; w = w^R\}$$

as follows

Case 1: $w = xx^R$

We evaluate

$w^R = (xx^R)^R = (x^R)^R x^R = xx^R = w$

We used property: $(x^R)^R = x$

Exercises

Case 2: $w = xax^R$

We evaluate

$w^R = (xax^R)^R = ((xa)x^R)^R = (x^R)^R(xa)^R = xax^R = w$

We used properties $(x^R)^R = x$ and $(xy)^R = y^R x^R$

Case 3: $w = xbx^R$

We evaluate

$w^R = (xbx^R)^R = ((xb)x^R)^R = (x^R)^R(xb)^R = xbx^R = w$

This **ends the proof**

Regular Grammars

Definition

A context-free grammar

$$G = (V, \ \Sigma, \ R, \ S)$$

is called **regular**, or **right-linear** if and only if

$$R \subseteq (V - \Sigma) \times \Sigma^*((V - \Sigma) \cup \{e\})$$

Regular Grammars

That is, a regular (right-linear) grammar is a context-free grammar such that the right-hand side of every **rule** contains at most one **nonterminal**, which if present, must be the last symbol in the string

The rules must have a form

$$A \to wB, \quad A \to w \quad \text{for any} \quad A, B \in V - \Sigma, \quad w \in \Sigma^*$$

Remark

We didn't say $A \neq B$!

Exercise 4

Given a **regular grammar** $G = (V, \Sigma, R, S)$, where

$V = \{a, b, S, A\}$, $\Sigma = \{a, b\}$

$$R = \{S \to aS \mid A \mid e, \ \ A \to abA \mid a \mid b\}$$

1. Construct a finite automaton M, such that $L(G) = L(M)$

Solution

We construct a non-deterministic finite automaton

$$M = (K, \ \Sigma, \ \Delta, \ s, \ F) \quad \text{for}$$

$$K = (V - \Sigma) \cup \{f\}, \ \Sigma = \Sigma, s = S, \ \ F = \{f\}$$

$$\Delta = \{(S, a, S), (S, e, A), (S, e, f), (A, ab, A), (A, a, f), (A, b, f)\}$$

Exercise 4

2. Write a computation of M that leads to the **acceptance** of the string *aaaababa*

Compare it with a derivation of the same string in G

Solution

The accepting computation of M is:

$$(S, aaaababa) \vdash_M (S, aaababa) \vdash_M (S, aababa) \vdash_M (S, ababa)$$

$$\vdash_M (A, ababa) \vdash_M (A, aba) \vdash_M (A, a) \vdash_M (f, e)$$

Corresponding **derivation** in G is:

$$S \Rightarrow aS \Rightarrow aaS \Rightarrow aaaS \Rightarrow aaaA \Rightarrow aaaabA$$

$$\Rightarrow aaaababA \Rightarrow aaaababa$$

Regular and Context-free Languages

We are going to prove the following **theorem** that establishes the **relationship** between the Regular Languages and Regular Grammars

L-G Theorem

Language L is **regular** if and only if there exists a **regular grammar** G such that

$$L = L(G)$$

Regular and Context-free Languages

By definition, any regular grammar is context free and hence generates a context-free language and we get that

R - CF Theorem

The the class RL of all regular languages is a proper subset of the class CFL of all context-free languages, i.e.

$$RL \subset CFL$$

Proof of L-G Theorem

L-G Theorem

Language L is **regular** if and only if there exists a **regular** grammar G such that

$$L = L(G)$$

Proof part 1

Suppose that L is **regular**; then L is accepted by a **deterministic** finite automaton

$$M = (K,\ \Sigma,\ \delta,\ s,\ F)$$

We **construct** a regular grammar G as follows

$$G = (V,\ \Sigma,\ R,\ S)$$

for $V = \Sigma \cup K, \quad S = s$

$$R = \{q \ \rightarrow \ ap : \quad \delta(q, a) = p\} \ \cup \ \{q \ \rightarrow \ e : \quad q \in F\}$$

Proof of L-G Theorem

We need now to show that L(M) = L(G)

Observe that the rules of G are designed to mimic exactly the moves of M

For any $\sigma_1, \ldots, \sigma_n \in \Sigma$ and $p_0, \ldots, p_n \in K$

$$(p_0, \sigma_1, \ldots, \sigma_n) \vdash_M (p_1, \sigma_2, \ldots, \sigma_n) \vdash_M \ldots \vdash_M (p_n, e)$$

if and only if

$$p_0 \overset{*}{\underset{G}{\Rightarrow}} \sigma_1 p_1 \overset{*}{\underset{G}{\Rightarrow}} \sigma_1\sigma_2 p_2 \ \ldots \ \overset{*}{\underset{G}{\Rightarrow}} \sigma_1\sigma_2 \ \ldots \ \sigma_n p_n$$

This is because

$$\delta(q, a) = p \quad \text{if and only if} \quad q \rightarrow ap$$

Proof of L-G Theorem

We **prove** now that $L(M) \subseteq L(G)$

Suppose that $w \in L(M)$

Then for some $p \in F$,

$$(s,\ w) \vdash_M{}^{*} (p, e)$$

but

$$\delta(q, a) = p \quad \text{if and only if} \quad q \ \rightarrow \ ap$$

is

$$S \ \overset{*}{\underset{G}{\Rightarrow}} \ w$$

so $w \in L(G)$

Proof of L-G Theorem

We **prove** now that $L(G) \subseteq L(M)$

Suppose that $w \in L(G)$

Then

$$S \overset{*}{\underset{G}{\Rightarrow}} w \quad \text{that is} \quad s \overset{*}{\underset{G}{\Rightarrow}} w$$

The rule **used** at the last step of the derivation must have been of the form

$$p \rightarrow e \quad \text{for some} \quad p \in F$$

and so

$$s \overset{*}{\underset{G}{\Rightarrow}} wp \underset{G}{\Rightarrow} w$$

But then

$$(s,\ w) \vdash_M{}^* (p, e)$$

and so $w \in L(M)$ and

$$L(M) = L(G)$$

Proof part 2

Let now G be any **regular** grammar

$$G = (V,\ \Sigma,\ R,\ S)$$

We define a **nondeterministic** automaton M such that

$$L(M) = L(G)$$

as follows

$$M = (K,\ \Sigma,\ \Delta,\ s,\ F)$$

$K = (V - \Sigma) \cup \{f\}$ where f is a new element

$$s = S, \qquad F = \{f\}$$

Proof of L-G Theorem

The set Δ of transitions is

$$\Delta = \{(A, w, B) : \;\; A \rightarrow wB \in R; \; A, B \in V - \Sigma, \;\; w \in \Sigma^*\}$$

$$\cup\{(A, w, f) : \;\; A \rightarrow w \in R; \; A, B \in V - \Sigma, \;\; w \in \Sigma^*\}$$

Once again, derivations are mimicked by the moves, i.e, for any

$A_1, \ldots, A_n \in V - \Sigma, \;\; w_1, \ldots w_n \in \Sigma^*$

$$A_1 \Rightarrow_G \; w_1 A_2 \; \Rightarrow_G \cdots \Rightarrow_G \; w_1 \ldots w_{n-1} A_n \Rightarrow_G \; w_1 \ldots w_n$$

if and only if

$$(A_1, w_1 \ldots w_n) \vdash_M (A_2, w_2 \ldots w_n) \vdash_M \; \ldots \; \vdash_M (A_n, w_n) \vdash_M (f, e)$$

Simple Questions

Justify if **True** or **False**

Q1 The set of terminals in a context free grammar G is a subset of the alphabet of G

Q2 The set of terminals and non- terminals in a context free grammar G form the alphabet of G

Q3 The set of non-terminals is always non- empty

Q4 The set of terminals is always non- empty

Simple Questions

Justify if **True** or **False**

Q1 The set of terminals in a context free grammar G is a subset of the alphabet of G

True By definition: $\Sigma \subseteq V$

Q2 The set of terminals and non- terminals in a context free grammar G form the alphabet of G

True By definition: $V = \Sigma \cup (V - \Sigma)$

Simple Questions

Justify if **True** or **False**

Q3 The set of non-terminals is always non- empty

True By definition: $S \in V$

Q4 The set of terminals is always non- empty

False $\Sigma = \emptyset$ is a finite set

Simple Questions

Justify if **True** or **False**

Q5 Let G be a context-free grammar

$$L(G) = \{w \in V : \quad S \underset{G}{\overset{*}{\Rightarrow}} w\}$$

Q6 The language $L \subseteq \Sigma^*$ is context-free if and only if

$$L = L(G)$$

Q7 A language is context-free if and only if it is accepted by a context-free grammar

Simple Questions

Justify if **True** or **False**

Q5 Let G be a context-free grammar

$$L(G) = \{w \in V : \quad S \underset{G}{\overset{*}{\Rightarrow}} w\}$$

False Should be $w \in \Sigma^*$

Q6 The language $L \subseteq \Sigma^*$ is context-free if and only if L = L(G)

False Holds only when G is a context -free grammar

Simple Questions

Justify if **True** or **False**

Q7 A language is context-free if and only if it is accepted by a context-free grammar

False Language is generated, not accepted by a grammar

Q8 Any regular language is context-free

True: Regular languages are generated by regular grammars, that are special case of CF grammars

Q9 Language is regular if and only if is generated by a regular grammar (right- linear)

True : Theorem proved in class

Simple Questions

Justify if **True** or **False**

Q10 $L = \{w \in \{a,b\}^* : \quad w = w^R\}$ is context-free

True: G with the rules:

$S \rightarrow aSa|bSb|a|b||e$ is the required grammar

Q11 A regular language is a CF language

True: Regular grammar is a special case of a context-free grammar