CSE 105

THEORY OF COMPUTATION

Spring 2018

http://cseweb.ucsd.edu/classes/sp18/cse105-ab/

Today's learning goals

Sipser Section 2.1

- Identify the components of a formal definition of a CFG
- Derive strings in the language of a given CFG
- Determine the language of a given CFG
- Design a CFG generating a given language

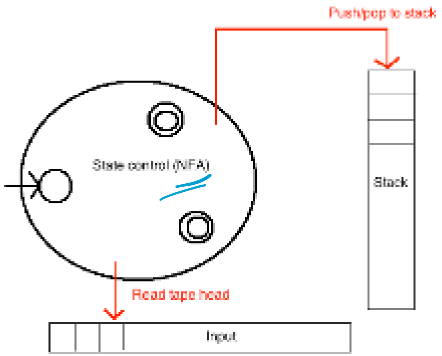
Mid quarter feedback: 1 participation point

https://goo.gl/forms/eo14OwIc84Bzjvda2

PDA: NFA as ??: RegExp

- Automata
  - String is read by the machine one character at a time, from left to right.
  - Determine if computation is successful by checking if entire string was read, and if land at an accept state.

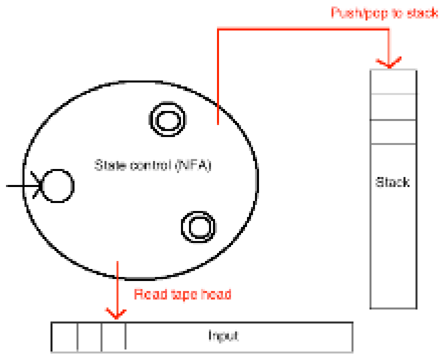
Ch1: Nondet didn't help



Note: PDA based on **N**FA; can't always be determinized.

PDA: NFA as ??: RegExp

- Automata
  - String is read by the machine one character at a time, from left to right.
  - Determine if computation is successful by checking if entire string was read, and if land at an accept state.
- Regular expressions and ??
  - Derive all strings in the language by following rules for required patterns.



Context-free grammar

Informally, a collection of rules used to *create* string.

CFGs *generate* languages.

Some sample rules:

$$S \to aTb$$
$$T \to aT$$
$$T \to bTS$$
$$S \to \varepsilon$$

LHS RHS

aTb

~~S~~

More formally...

Context-free grammar

Sipser Def 2.2, page 102

(V, Σ, R, S)

Variables: finite set of (usually upper case) variables **V**

Terminals: finite set of alphabet symbols **Σ** $V \cap \Sigma = \emptyset$

Rules/Productions: finite set of allowed transformations **R**

$A \rightarrow u$ $A \in V, u \in (V \cup \Sigma)^*$

LHS RHS

Start variable: origination of each derivation **S**

Derivation

Set of variables

Set of terminals

$$G = (\{S\}, \{0\}, R, S)$$

with the rules

$$R = \{S \xrightarrow{1} 0S,\ S \xrightarrow{2} 0\} \rightarrow$$

Sample derivation:

$$S \overset{1}{\Rightarrow} 0S \overset{1}{\Rightarrow} 00S \overset{2}{\Rightarrow} 000 \quad \Rightarrow$$

Start variable

One-step application of rule

String of terminals

$000 \in L(G)$

Context-free language

Sipser p. 104

The **language generated by a CFG** (V, Σ, R, S) is

{ w in Σ^* | starting with the Start variable and applying sequence of rules, can derive w on RHS}

If G = (V, Σ, R, S) the language generated by G is denoted L(G).

Notation:

$S \Longrightarrow^* w$

Terminology: sequence of rule applications is **derivation**

Context-free language

Sipser p. 104

The language generated by CFG (V, Σ, R, S) is

{ w in Σ^* | starting with the Start variable and applying sequence of rules, can derive w on RHS}.

What is the language generated by the CFG ({S}, {0}, R, S) with the rules R = {S → 0S, S → 0} ?

~~A.~~ {0}

~~B.~~ {0, 0S}

C. {0, 00, 000, …} $= L(00^*)$

~~D.~~ {ε, 0, 00, 000, …}

E. I don't know.

000

Context-free language

Sipser p. 104

What is the language of the CFG ({S}, {0,1}, R, S) with rules

S → 0S ①
S → 1S ②
S → ε ③

LHS

S → 0S | 1S | ε

RHS₁ RHS₂ RHS₃

A. L(0*1*)
B. L(0* U 1*)
C. L((0 U 1)*)
D. L ((0*1*)*)
E. I don't know.

ε

0*1*

11 = ε 11

10 = ε 1 0 ε

010

0/1 0/1 0/1

01 0ε

0*1* 0*1*

Designing a CFG

L = { abba }

Which CFG generates L?

$S \Rightarrow aT \Rightarrow abV \Rightarrow abbW \Rightarrow abba$

A. ({S, T, V, W} , {a, b}, {S → aT , T → bV , V → bW , W → a }, S)
B. ({Q}, {a,b}, { Q → abba } , Q)
C. ({X, Y} , {a,b}, { X → aYa , Y → bb } , X)
D. All of the above
E. None of the above

$Q \Rightarrow abba$

$X \Rightarrow aYa \Rightarrow abba$

Context-free languages

- L (00*)
- L((0U1)*)
- { abba } + any finite set

Is any nonregular set context-free?

What about the languages that are recognized by PDAs?

Designing a CFG

We know this set is not regular!

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

(V, Σ, R, S)

$\varepsilon \in L$

$S \to aSb \mid \varepsilon$

aaabbb

i.e.

$(\{S\}, \{a, b\}, \{S \to aSb \mid \varepsilon\}, S)$

Designing a CFG

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

One approach:

- what is shortest string in the language?
- how do we go from shorter strings to longer ones?

Context-free languages

- L (00*)
- L((0U1)*)
- { abba }
- $\{ a^n b^n \mid n \geq 0 \}$

$S \to \varepsilon \mid 0S1 \mid S1$

Ex: $\{ \underline{0^i 1^j} \mid j \geq i \geq 0 \}$ recognizable by a PDA

$0^i 1^i 1^{j-i}$

$S \to 0S1 \mid T$

$T \to 1T \mid \varepsilon$

PDAs and CFGs are equally expressive

Theorem 2.20: A language is context-free if and only if some nondeterministic PDA recognizes it.

Consequences

- Quick proof that every regular language is context free
- To prove closure of class of CFLs under a given operation, can choose two modes of proof (via CFGs or PDAs) depending on which is easier