

Context-Free Grammars

Introduction

Two views on grammars:

1. Grammars are rules defining the structure between symbols and clauses in a language
2. Grammars define mathematical transformation between symbols and clauses in a language

Using Grammars: generation and parsing

- Generation: using the grammar to produce strings in the language
- Parsing: using the grammar to accept or reject strings in the language, and also often to identify the structure in the string

Grammars: recursive representation of structure of a formal language

Outline

- Finite Automata and Grammars
- Context-Free Grammars
- Natural Language Grammars

Finite Automata and Grammars

Edge set: The edge set E for a DFA is $E = \{(q_i, \sigma, q_j) \mid q_j = \delta(q_i, \sigma)\}$

This is essentially the edge parameter for `make-fa` in project 1

For accept states, add edge (q_a, ϵ) where q_a is an accepting state

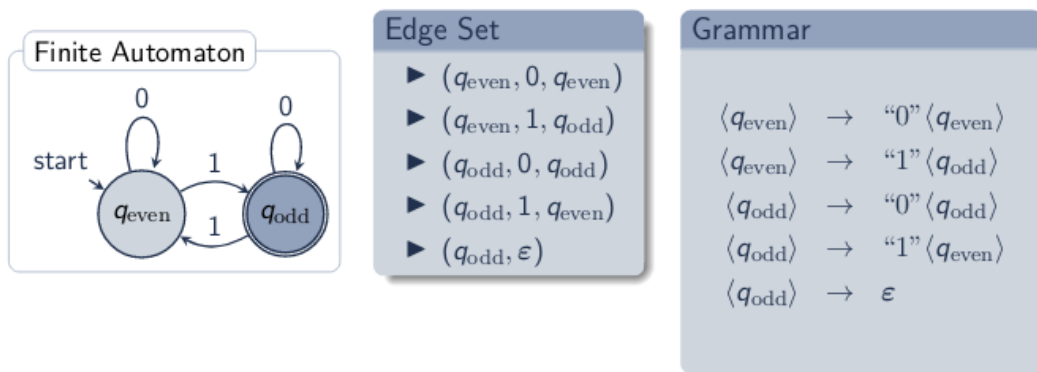


Figure 1: Example Finite Automata and its respective edge set and grammar

Backus-Naur Form

Using example from figure 1: here is the respective BNF grammar:

$$\langle q_{\text{even}} \rangle \rightarrow "0" \langle q_{\text{even}} \rangle, "1" \langle q_{\text{odd}} \rangle$$

$$\langle q_{\text{odd}} \rangle \rightarrow "0" \langle q_{\text{odd}} \rangle, "1" \langle q_{\text{even}} \rangle, \epsilon$$

Grammar Parts

- Terminal: Atomic, input symbol (Ex. "0")
- Nonterminal: Can replace with multiple elements (Ex. $\langle q_{\text{odd}} \rangle$)
- Production: Rewrite or substitution rule (Ex. $\langle q_{\text{even}} \rangle \rightarrow "0" \langle q_{\text{even}} \rangle$)

Generation Using Grammars

1. Begin: with a sequence containing starting nonterminal
2. Repeat: replace some nonterminal in the current sequence with a right-hand side for that nonterminal
3. Terminate: when no nonterminal are left in the sequence, i.e., we have a string of terminals

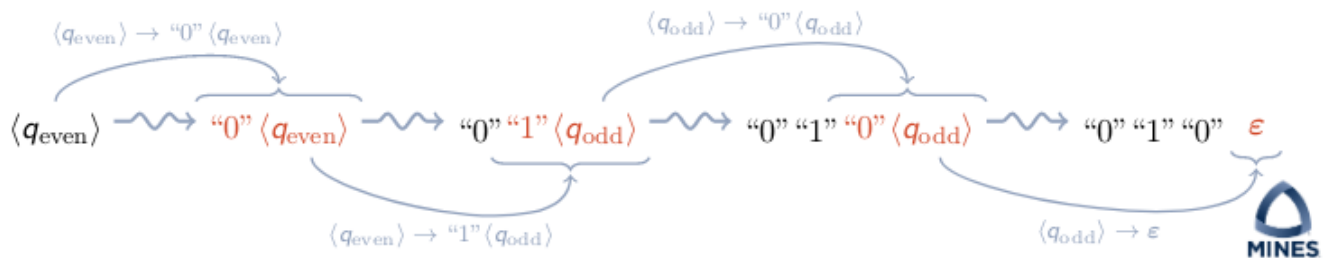
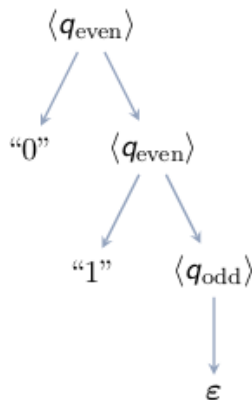


Figure 2: Example of generation using the grammar from figure 1

Parse Trees

- Leaves: Terminals
- Interior Nodes: Nonterminals
- Edges: Rewrite (Replacement) of a nonterminal



If we added an edge from 1 state to itself (plus quotations), we would have infinite grammar and infinite trees. This could cause our parse tree to create infinite quotations. We can ignore this with context free grammar.

Context-Free Grammar

Context-Free Grammar: A context-free grammar G is the tuple $G = (V, T, P, S)$, where:

- V is a finite set of variables or nonterminals
- T is a finite set of terminals
- P is a finite set of productions in the form $V \rightarrow (V \cup T)^*$
- $S \in V$ is the start symbol

Example

Grammar:

$$\langle e \rangle \rightarrow "(\langle e \rangle)", \epsilon$$

Corresponding Context-Free Grammar:

- $V = \{\langle e \rangle\}$
- $T = \{ "(", ")" \}$
- $P = \{ (\langle e \rangle \rightarrow "(\langle e \rangle)"), (\langle e \rangle \rightarrow \epsilon) \}$
- $S = \{\langle e \rangle\}$

Context-Free Grammar: Language of a Grammar

The language of CFG $G = (V, T, P, S)$, denoted $L(G)$, is the set of terminal strings that have derivations from the start symbol S ,

$$L(G) = \{ \omega \in T^* \mid S \rightsquigarrow \omega \}$$

What is “Context Free”

Rewriting nonterminals is independent of surrounding context. Meaning when we expand a nonterminal in the CFG, it won't duplicate unnecessary terminals.

Example: Infix Arithmetic

$$T = \{ +, -, *, \div, \text{number}, \text{variable}, (,) \}$$

Grammar:

- $\langle e \rangle \rightarrow \langle e \rangle \langle \text{op} \rangle \langle e \rangle \mid "(\langle e \rangle)" \mid \text{"number"} \mid \text{"variable"}$
- $\langle \text{op} \rangle \rightarrow + \mid - \mid * \mid \div$

Issue with this grammar:

If we try to construct a parse tree for $a + b * c$. We could have two representations. This is called an **ambiguous grammar**

Natural Language Grammars

Grammar for English subset:

$\langle \text{sentence} \rangle \rightarrow \langle \text{noun-phrase} \rangle \langle \text{verb-phase} \rangle$

$\langle \text{noun-phrase} \rangle \rightarrow \langle \text{noun} \rangle \mid \langle \text{adjective} \rangle \langle \text{noun-phrase} \rangle \mid \langle \text{article} \rangle \langle \text{noun-phrase} \rangle$

$\langle \text{verb-phase} \rangle \rightarrow \langle \text{verb} \rangle \mid \langle \text{verb} \rangle \langle \text{adverb} \rangle$