CSE 311 Foundations of Computing I

Lecture 19
Recursive Definitions:
Context-Free Grammars and Languages
Spring 2013

Announcements

- Reading assignments
  - 7th Edition, pp. 851-855
  - 6th Edition, pp. 789-793
- Today and Friday
  - 7th Edition, Section 9.1 and pp. 594-601
  - 6th Edition, Section 8.1 and pp. 541-548

Highlights … Languages: Sets of Strings

- Sets of strings that satisfy special properties are called *languages*. Examples:
  - English sentences
  - Syntactically correct Java/C/C++ programs
  - $\Sigma^* =$ All strings over alphabet Σ
  - Palindromes over Σ
  - Binary strings that don't have a 0 after a 1
  - Legal variable names. keywords in Java/C/C++
  - Binary strings with an equal # of 0's and 1's

Highlights…Regular expressions

- Regular expressions over Σ
- Basis:
  - $\varnothing$, λ are regular expressions
  - ***a*** is a regular expression for any $a \in \Sigma$
- Recursive step:
  - If **A** and **B** are regular expressions then so are:
    - (**A** $\cup$ **B**)
    - (**AB**)
    - **A***

Fact: Not all sets of strings can be specified by regular expressions

- Even some easy things like
  - Palindromes
  - Strings with equal number of 0's and 1's
- But also more complicated structures in programming languages
  - Matched parentheses
  - Properly formed arithmetic expressions
  - Etc.

Context Free Grammars

- A Context-Free Grammar (CFG) is given by a finite set of substitution rules involving
  - A finite set **V** of *variables* that can be replaced
  - Alphabet Σ of *terminal symbols* that can't be replaced
  - One variable, usually **S**, is called the *start symbol*
- The rules involving a variable **A** are written as

  $\mathbf{A} \rightarrow w_1 \mid w_2 \mid \ldots \mid w_k$ where each w_i is a string of variables and terminals – that is $w_i \in (\mathbf{V} \cup \Sigma)^*$

How Context-Free Grammars generate strings

- Begin with start symbol **S**
- If there is some variable **A** in the current string you can replace it by one of the w's in the rules for **A**
  - Write this as $x\mathbf{A}y \Rightarrow xwy$
  - Repeat until no variables left
- The set of strings the CFG generates are all strings produced in this way that have no variables

Sample Context-Free Grammars

- Example: $\mathbf{S} \rightarrow 0\mathbf{S}0 \mid 1\mathbf{S}1 \mid 0 \mid 1 \mid \lambda$

- Example: $\mathbf{S} \rightarrow 0\mathbf{S} \mid \mathbf{S}1 \mid \lambda$

Sample Context-Free Grammars

- Grammar for $\{0^n1^n : n \geq 0\}$ all strings with same # of 0's and 1's with all 0's before 1's.
- Example: $\mathbf{S} \rightarrow (\mathbf{S}) \mid \mathbf{SS} \mid \lambda$

Simple Arithmetic Expressions

$\mathbf{E} \rightarrow \mathbf{E}+\mathbf{E} \mid \mathbf{E}*\mathbf{E} \mid (\mathbf{E}) \mid x \mid y \mid z \mid 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

Generate $(2*x) + y$

Generate $x+y*z$ in two fundamentally different ways

Context-Free Grammars and recursively-defined sets of strings

- A CFG with the start symbol **S** as its only variable recursively defines the set of strings of terminals that **S** can generate
- A CFG with more than one variable is a simultaneous recursive definition of the sets of strings generated by *each* of its variables
  - Sometimes necessary to use more than one

Building in Precedence in Simple Arithmetic Expressions

- **E** – expression (start symbol)
- **T** – term **F** – factor **I** – identifier **N** - number

$\mathbf{E} \rightarrow \mathbf{T} \mid \mathbf{E}+\mathbf{T}$

$\mathbf{T} \rightarrow \mathbf{F} \mid \mathbf{F}*\mathbf{T}$

$\mathbf{F} \rightarrow (\mathbf{E}) \mid \mathbf{I} \mid \mathbf{N}$

$\mathbf{I} \rightarrow x \mid y \mid z$

$\mathbf{N} \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

Another name for CFGs

- BNF (Backus-Naur Form) grammars
  - Originally used to define programming languages
  - Variables denoted by long names in angle brackets, e.g.
    - <identifier>, <if-then-else-statement>, <assignment-statement>, <condition>
    - ::= used instead of $\rightarrow$

BNF for C

```
statement:
  ((identifier | "case" constant-expression | "default") ":")*
  (expression? ";" |
   block |
   "if" "(" expression ")" statement |
   "if" "(" expression ")" statement "else" statement |
   "switch" "(" expression ")" statement |
   "while" "(" expression ")" statement |
   "do" statement "while" "(" expression ")" ";" |
   "for" "(" expression? ";" expression? ";" expression? ")" statement |
   "goto" identifier ";" |
   "continue" ";" |
   "break" ";" |
   "return" expression? ";"
  )

block: "{" declaration* statement* "}"

expression:
  assignment-expression%

assignment-expression: (
    unary-expression (
      "=" | "*=" | "/=" | "%=" | "+=" | "-=" | "<<=" | ">>=" | "&=" |
      "^=" | "|="
    )
  )* conditional-expression

conditional-expression:
  logical-OR-expression ( "?" expression ":" conditional-expression )?
```


Parse Trees

Back to middle school:

<sentence>::=<noun phrase><verb phrase>

<noun phrase>::=<article><adjective><noun>

<verb phrase>::=<verb><adverb>|<verb><object>

<object>::=<noun phrase>

Parse:

The yellow duck squeaked loudly

The red truck hit a parked car

CSE 311 Foundations of Computing I

Lecture 19 continued

Relations

Spring 2013

Definition of Relations

Let A and B be sets,
A binary relation from A to B is a subset of $A \times B$

Let A be a set,
A binary relation on A is a subset of $A \times A$

Relation Examples

$R_1 = \{(a, 1), (a, 2), (b, 1), (b, 3), (c, 3)\}$

$R_2 = \{(x, y) \mid x \equiv y \pmod 5\}$

$R_3 = \{(c_1, c_2) \mid c_1$ is a prerequisite of $c_2\}$

$R_4 = \{(s, c) \mid$ student s has taken course c $\}$

Properties of Relations

Let R be a relation on A

R is reflexive iff $(a,a) \in R$ for every $a \in A$

R is symmetric iff $(a,b) \in R$ implies $(b, a) \in R$

R is antisymmetric iff $(a,b) \in R$ and $a \neq b$ implies $(b,a) \notin R$

R is transitive iff $(a,b) \in R$ and $(b, c) \in R$ implies $(a, c) \in R$