Compsci 334 - Formal Languages and Automata with Applications
Dr. Susan Rodger
Section: Introduction (Ch. 1) (handout)

Announcements:

- This is a math course with systems applications. Prereq: CompSci 201, Compsci 230 or equiv.

- Course web page:

  www.cs.duke.edu/courses/spring26/compsci334

  Familiarize yourself with all parts of the web page. All lecture notes, assignments and resources will be on this page.

- Flipped class - you have to complete reading and quizzes BEFORE you come to class

- Read Chapter 1 in the Linz/Rodger book for next time. (Readings posted on website)

- Complete the reading quizzes on Canvas before class. We will have reading quizzes due almost every class period. They must be completed before the next class period starts, by 11:45am, and will be based on assigned reading for that date. They turn off automatically when class starts. Canvas will be used mostly for reading quizzes and also for posting grades.

  (Due to Drop/add, QZ01-QZ05 turn off Jan. 22, 11:45am!!, You should try to do them before class starts to get in the habit of doing them on time! After Jan. 22, a quiz turns off when class starts and we don't turn them back on. You will have a week to do them.)

- Course bulletin board: Ed Discussion (get to from Canvas)

- Course participation required!

What will we do in Compsci 334?

Questions

- Can you write a program to determine if a string is an integer?

9998.89	8abab	789342

- Can you do this if your machine had no additional memory other than the program? (can't store any values and look at them again)

- Can you write a program to determine if the following are correct arithmetic expressions?
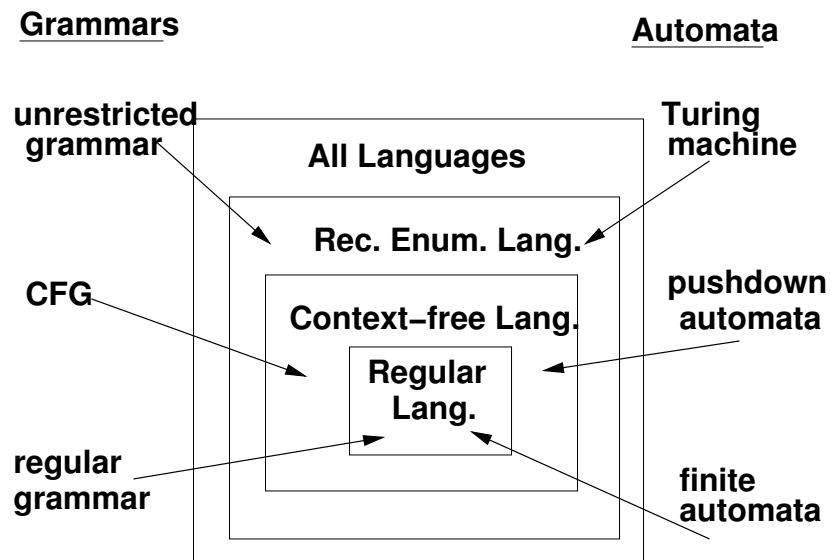
$$((34 + 7 * (18/6)))$$

$$(((((((a + b) + c) * d(e + f)))))$$

- Can you do this if your machine had no additional memory other than the program?

- Can you write a program to determine the **value of** the following expression?

$$((34 + 7 * (18/6)))$$

- Can you write a program to determine if a file is a valid Java program?

- Can you write a program to determine if a Java program given as input will ever halt?
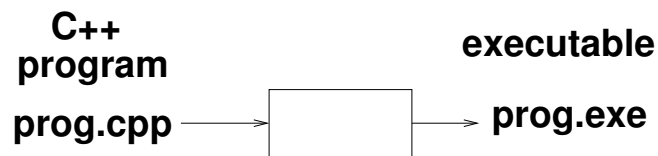
Language Hierarchy

Power of Machines

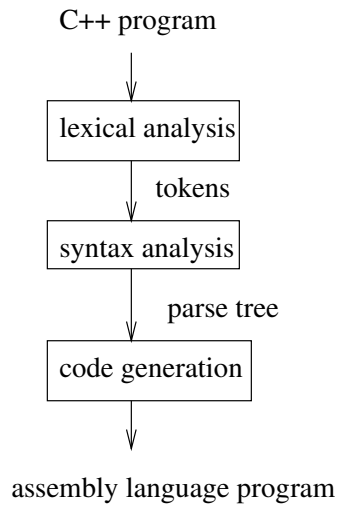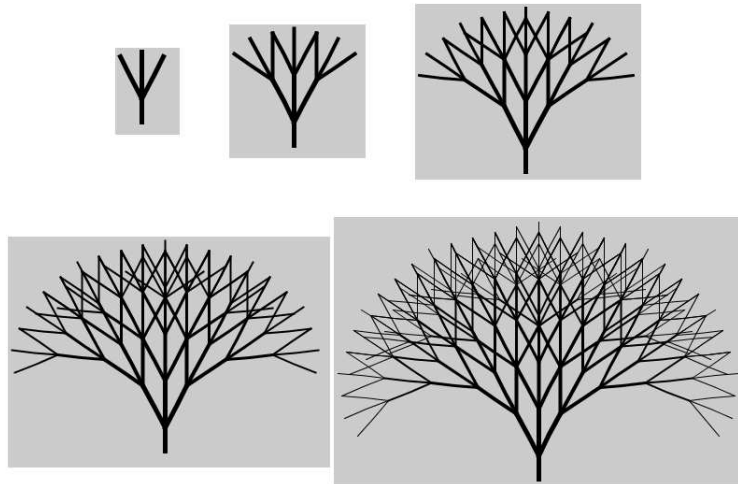| automata	Can do?	Can't do?
FA (no memory)		
PDA (stack)		
TM (infinite)		

Application

Compiler

- Our focus - Question: Given a program in some language (say Java or C++) - is it valid?

- Question: language L, program P - is P valid?

- Other things to consider, how is the compilation process different for different programming languages? (Java vs C++?)

2

Stages of a Compiler

C++ program

$\downarrow$

| lexical analysis |

tokens $\downarrow$

| syntax analysis |

parse tree $\downarrow$

| code generation |

$\downarrow$

assembly language program

L-Systems - Model the Growth of Plants

Set Theory - Read Chapter 1 Linz/Rodger.

A Set is a collection of elements.

A={1,4,6,8}, B={2,4,8}, C={3,6,9,12,...}, D={4,8,12,16,...}

- (union) A∪B=
- (intersection) A∩B=
- C∩D=
- (member of) $42 \in$ C?
- (subset) B⊂C?
- B∩A ⊆D?
- |B|=

- (product) A×B=

- |A×B|=

- ∅ ∈ B∩C?

- (powerset) $2^B=$

Example What are all the subsets of $\{3, 5\}$?

How many subsets does a set S have?

$|S|$	number of subsets
0	
1	
2	
3	
4	

How do you prove? Set S has $2^{|S|}$ subsets.

Technique: Proof by Induction

1. Basis: P(1)? Prove smallest instance is true.

2. Induction Hypothesis - I.H.
   Assume P(n) is true for 1,2,...,n

3. Induction Step - I.S.
   Show P(n+1) is true (using I.H.)

Set S has $2^{|S|}$ subsets.

Proof:

1. Basis:

2. I.H. Assume

3. I.S. Show

Ch. 1: 3 Major Concepts

- languages
- grammars
- automata

Languages

- Σ - set of symbols, alphabet
- string - finite sequence of symbols
- language - set of strings defined over Σ

alphabet Σ

Examples

- $\Sigma = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
  L=$\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, ...\}$
- $\Sigma = \{a, b, c\}$
  L=$\{ab, ac, cabb\}$
- $\Sigma = \{a, b\}$
  L=$\{a^n b^n \mid n > 0\}$

Notation

- symbols in alphabet: a, b, c, d, ...
- string names: u,v,w,...

Definition of concatenation

Let w=$a_1 a_2 \ldots a_n$ and v=$b_1 b_2 \ldots b_m$

Then $w \circ v$ OR wv=

See book for formal definitions of other operations.

String Operations

strings: w=abbc, v=ab, u=c

- size of string
  $|w| + |v| =$

- concatenation

  $v^3 = \text{vvv} = \text{vovov} =$

- $v^0 =$

- $w^R =$

- $|vv^R w| =$

- ab $\circ$ $\lambda =$

Definition

$\Sigma^* =$ set of strings obtained by concatenating 0 or more symbols from Σ

Example

$\Sigma = \{a, b\}$

$\Sigma^* =$

$\Sigma^+ =$

Examples

$\Sigma = \{a, b, c\}, \ L_1 = \{ab, bc, aba\}, \ L_2 = \{c, bc, bcc\}$

- $L_1 \cup L_2 =$

- $L_1 \cap L_2 =$

- $\overline{L_1} =$

- $\overline{L_1 \cap L_2} =$

- $L_1 \circ L_2 = \{xy \mid x \in L_1 \text{ and } y \in L_2\} =$

Definition

$L^0 = \{\lambda\}$

$L^2 = L \circ L$

$L^3 = L \circ L \circ L$

$L^* = L^0 \cup L^1 \cup L^2 \cup L^3 \ldots$

$L^+ = L^1 \cup L^2 \cup L^3 \ldots$

Grammars

Grammar for english

\<sentence\> → \<subject\>\<verb\>\<d.o.\>

\<subject\> → \<noun\> | \<article\>\<noun\>

\<verb\> → hit | ran | ate

\<d.o.\> → \<article\>\<noun\> | \<noun\>

\<noun\> → Fritz | ball

\<article\> → the | an | a

Examples (derive a sentence)

Fritz hit the ball.

```
<sentence> -> <subject><verb><d.o>
            -> <noun><verb><d.o>
            -> Fritz <verb><d.o.>
            -> Fritz hit <d.o.>
            -> Fritz hit <article><noun>
            -> Fritz hit the <noun>
            -> Fritz hit the ball
```

Can we also derive the sentences?

```
The ball hit Fritz.
```

```
The ball ate the ball
```

Syntactically correct?

Semantically correct?

Grammar

G=(V,T,S,P) where

- V - variables (or nonterminals)
- T - terminals
- S - start variable (S∈V)
- P - productions (rules)
  x→y "means" replace x by y
  x∈(V∪ T)$^{+}$, y∈(V∪T)*
  where V, T, and P are finite sets.

Definition

w $\Rightarrow$ z	w derives z
w $\overset{*}{\Rightarrow}$ z	derives in 0 or more steps
w $\overset{+}{\Rightarrow}$ z	derives in 1 or more steps

Definition of Language of a grammar - L(G)

G=(V,T,S,P)

L(G)={w$\in T^*$ | S $\overset{*}{\Rightarrow}$ w}

Example

G=({S}, {a,b}, S, P)

P={S$\to$ aaS, S$\to$ b}

L(G)=

Example

L(G) = $\{a^n ccb^n \mid n > 0\}$

G =

Example

G=({S}, {a,b}, S, P)

P={S$\to$ aSb, S$\to$ SS, S$\to$ ab}

Which of these strings $aabb, abab, abba, babab$ can be generated by this grammar? Show the derivations.

L(G) =

Automata Abstract model of a digital computer