COMS W3261
Computer Science Theory
Lecture 5
Properties of Regular Expressions

Alexander Roth

2014 – 09 – 17

Overview

- Algebraic laws can be used to simplify regular expressions.
- The pumping lemma for regular languages can be used to prove that some languages are not regular.
- The set of regular languages is closed under many common operations such as union, intersection, complement and reversal.

1 Algebraic Laws for Regular Expressions

- Algebraic laws can be used to simplify regular expressions
- Here are some of the most important algebraic identities for regular expressions
  - Union is commutative: $L + M = M + L$
  - Union is associative: $(L + M) + N = L + (M + N)$
  - Concatenation is associative: $(LM)N = L(MN)$
  - $\varnothing$ is the identity for union: $\varnothing + L = L + \varnothing = L$
  - ϵ is the identity for concatenation: $\epsilon L = L\epsilon = L$
  - $\varnothing$ is the annihilator for concatenation: $\varnothing L = L\varnothing = \varnothing$
  - Concatenation left distributes over union: $L(M + N) = LM + LN$
  - Concatenation right distributes over union: $(M + N)L = ML + NL$
  - Union is idempotent: $L + L = L$
  - $L^{**} = L^*$
  - $\varnothing^* = \epsilon$
  - $\epsilon^* = \epsilon$

1.1 Class Notes

Simplifying regular expressions:

$$\begin{aligned}
&\epsilon + a + (\epsilon + a)(\epsilon + a)^*(\epsilon + a) \\
&\epsilon + a + (\epsilon + a)a^*(\epsilon + a) \\
&\epsilon + a + a^* \\
&a^*
\end{aligned}$$

2 The Pumping Lemma for Regular Languages

- The pumping lemma for regular languages states that for every nonfinite regular language L, there exists a constant n that depends on L such that for all w in L with $|w| \geq n$, there exists a decomposition of w into xyz such that

  1. $y \neq \epsilon$
  2. $|xy| \leq n$, and
  3. for all $k \geq 0$, the string xy^kz is in L.

- Proof: See HMU, p. 129.

- One important use of the pumping lemma is to prove some languages are not regular.

- Example: The language L consisting of all strings of a's and b's of the form a^ib^i, $i \geq 0$, is not regular.

  - The proof will be by contradiction. Assume L is regular. Then by the pumping lemma there is a constant n associated with L such that for all w in L with $|w| \geq n$, w can be written as xyz such that

    1. $y \neq \epsilon$
    2. $|xy| \leq n$, and
    3. for all $k \geq 0$, the string xy^kz is in L.

  - Since $|xy| \leq n$, $xy = a^m$ for some $0 < m \leq n$.
  - Setting $k = 0$, condition (3) of the pumping lemma says xz must also be in L.
  - But xz is of the form a^pb^n, where $p < n$.
  - This contradicts the conclusion that xz must be in L.

3 Closure Properties of Regular Languages

- A closure property for a family of languages is a theorem that says if we apply a certain operation to the languages in the family, then the resulting language will also be in the family. For example. if we take the union of two regular languages L and M, then the language $L \cup M$ is also regular. We therefore say the regular languages are closed under the operation of union.
- We can show that the regular languages are closed under the following operations:
  - union, intersection, complement, difference
  - concatenation, Kleene closure
  - reversal
  - homomorphism, inverse homomorphism
- These closure properties can be used to show that some languages are regular.
- These closure properties combined with the pumping lemma can be used to show some languages are not regular.

4 Practice Problems

1. Show that the language consisting of all strings of balanced parentheses is not regular.
2. Prove that the language consisting of all strings of a's and b's that read the same forwards as backwards is not regular.
3. Prove that the language $L = \{w \,|\, w = a^i b^i$ where i is not equal to $j\}$ is not regular.

5 Reading Assignment

- HMU: Sects. 3.4, 4.1, 4.2