Nonregular languages

Sipser 1.4 (pages 77-82)

Nonregular languages?

• We now know:
  – Regular languages may be specified either by
    • regular expressions
    • DFAs or NFAs
  • What if we can’t find a regular expression or finite state automaton for a language?
  • How do we show a language is not regular?
Limited memory

• Since finite state automata cannot back up when reading an input, they are allowed only a bounded amount of memory
• What about the language
  \( \{0^n 1^n \mid n \geq 0\} \)?

Hmm... how can we prove a language is not regular?

• What about
  – \( \{w \mid w \text{ has an equal number of } 0\text{s and } 1\text{s}\} \)
  – \( \{w \mid w \text{ has an equal number of occurrences of the substrings } 01 \text{ and } 10\} \)
Try a different perspective

- How can a regular language be infinite?
- Iff its regular expression must have a Kleene star
- Star operators correspond to cycles in the finite state automaton

Pigeonhole principle

- Let $M$ be a finite state machine with $N$ states recognizing an infinite language
- Let $x \in L(M)$ with $|x| = N$
- Then there exists a sequence of states $S_0, S_1, S_2, \ldots, S_N$
- So: $N+1$ pigeons into $N$ holes...
  - Some hole must have at least 2 pigeons!
  - I.e., at least two of the states must be the same, so there must be a cycle
Machine loops

• Let $q_k$ be the first repeated state; that is, $q_k = q_{k+p}$ for some $p$, $0 \leq k < k+p \leq N$.

• Where

  $x = a_1a_2...a_k...a_{k+p}...a_N = uvw$
  $u = a_1a_2...a_k$
  $v = a_{k+1}...a_{k+p}$
  $w = a_{k+p+1}...a_N$

• We conclude: $uv^i w \in L(M)$ for all $i \geq 0$.

The pumping lemma

• Theorem 1.70: If $A$ is a regular language, then there is a number $N$ where, if $x$ is any string of length at least $N$, then $x = uvw$, such that

  1. For each $i \geq 0$, $uv^i w \in A$,
  2. $|v| > 0$, and
  3. $|uv| \leq N$. 
So now...

- Is \( L = \{0^n1^n : n \geq 0\} \) regular?
- Prove it!
- If \( L \) is regular... we can apply the pumping lemma!
  
  ...there are strings \( u, v, \) and \( w \) such that 
  
  \[ uv^i w \in L \text{ for all } i \geq 0. \]
  
  – What does \( v \) look like?
    - Entirely 0s
    - Entirely 1s
    - Both 0s and 1s

Reuse!

- Is \( C = \{ w \mid w \text{ has an equal number of 0s and 1s} \} \) a regular language?
Picking the substring to pump

• Is $PAL = \{w \in \{0, 1\}^* : w \text{ is a palindrome}\}$ a regular language?