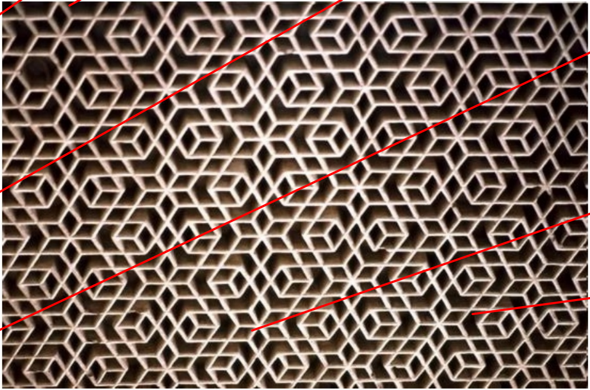
Automata, Languages and Computation

Chapter 4 : Properties of Regular Languages

Master Degree in Computer Engineering
University of Padua
Lecturer : Giorgio Satta

Lecture based on material originally developed by :
Gösta Grahne, Concordia University

Properties of regular languages



1. Pumping Lemma : every regular language satisfies this property; useful to show that some languages are not regular
2. Closure properties : how to combine automata using specific operations
3. Decision problems : algorithms for the solution of problems based on automata/regex and their complexity
4. Automata minimization : reduce number of states to a minimum

Introduction to pumping lemma

Suppose $L_{01} = \{0^n 1^n \mid n \geqslant 1\}$ were a regular language

Then L_{01} must be recognized by some DFA A; let k be the number of states of A

Assume A reads 0^k. Then A must go through the following transitions :

$$\begin{array}{ll} \epsilon & p_0 \\ 0 & p_1 \\ 00 & p_2 \\ \dots & \dots \\ 0^k & p_k \end{array}$$

By the **pigeonhole principle**, there must exist a pair i, j with $i < j \leqslant k$ such that $p_i = p_j$. Let us call q this state

Introduction to pumping lemma

Now you can **fool** A :

- if $\hat{\delta}(q, 1^i) \notin F$, then the machine will foolishly reject $0^i 1^i$
- if $\hat{\delta}(q, 1^i) \in F$, then the machine will foolishly accept $0^j 1^i$

In other words: state q would represent inconsistent information about the count of occurrences of 0 in the string read so far

Therefore A does not exists, and L_{01} is not a regular language

Pumping lemma for regular languages

Theorem Let L be any regular language. Then $\exists n \in \mathbb{N}$ depending on L, $\forall w \in L$ with $|w| \geqslant n$, we can factorize $w = xyz$ with :

- $y \neq \epsilon$
- $|xy| \leqslant n$
- $\forall k \geqslant 0,\ xy^k z \in L$

Pumping lemma for regular languages

Proof

Suppose L is a regular language

Then L is recognized by some DFA A with, say, n states

Let $w = a_1 a_2 \cdots a_m \in L$ with $m \geqslant n$

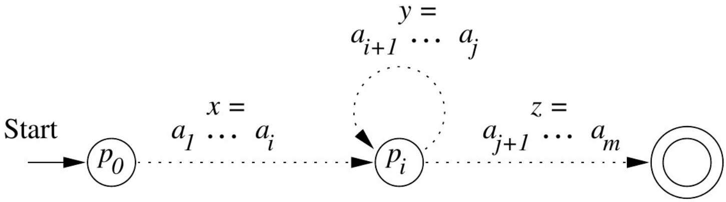
Let $p_i = \hat{\delta}(q_0, a_1 a_2 \cdots a_i)$, for each $i = 0, 1, \ldots, n$

There exists $i < j \leqslant n$ such that $p_i = p_j$

Pumping lemma for regular languages

Let us write $w = xyz$, where

- $x = a_1 a_2 \cdots a_i$
- $y = a_{i+1} a_{i+2} \cdots a_j$
- $z = a_{j+1} a_{j+2} \ldots a_m$



Evidently, $xy^k z \in L$, for any $k \geqslant 0$ □

Example

Let Σ be some alphabet, and let $w \in \Sigma^*$, $a \in \Sigma$. We write $\#_a(w)$ to denote the **number of occurrences** of a in w

We define

$$L_{eq} = \{w \mid w \in \{0,1\}^*, \#_0(w) = \#_1(w)\}$$

In words, L_{eq} is the language whose strings have an equal number of 0's and 1's

Use the pumping lemma to show that L is not regular

Example

Proof Suppose L_{eq} were regular. Then $L(A) = L_{eq}$ for some DFA A

Let n be the number of states of A and let $w = 0^n1^n \in L(A)$

By the pumping lemma we can factorize $w = xyz$ with

- $|xy| \leqslant n$,
- $y \neq \epsilon$

and state that, for each $k \geqslant 0$, we have $xy^kz \in L(A)$

$$w = \underbrace{000\cdots}_{x}\underbrace{\cdots 00}_{y}\underbrace{\cdots 0111\cdots 11}_{z}$$

Example

Proof (alternative) We can see the application of the pumping lemma as a game between two players

Player P2 states that L_{eq} is regular, and player P1 wants to establish a **contradiction**

- P2 picks n (number of states of DFA, if it exists)
- P1 picks string $w = 0^n 1^n \in L_{eq}$, with $|w| \geqslant n$
- P2 picks a factorization $w = xyz$, with $|xy| \leqslant n$, $y \neq \epsilon$ and $xy^k z \in L_{eq}$ (assuming L_{eq} is regular)
- P1 picks k such that $xy^k z \notin L$, which is a violation of the pumping lemma. Specifically, P1 picks $k = 0$: $xz \notin L_{eq}$, since y contains just 0's, $y \neq \epsilon$, and thus $\#_0(xz) < \#_1(xz) = n$
- P1 concludes that L_{eq} cannot be regular $\square$

Example

Let $L_{pr} = \{1^p \mid p \text{ prime}\}$. Using the pumping lemma, show that L_{pr} is not regular

Proof Let n be as in the pumping lemma, and let $p \geqslant n+2$ be some prime number. Thus $1^p \in L_{pr}$

By the pumping lemma we can write $w = xyz$ with

- $|xy| \leqslant n$,
- $y \neq \epsilon$

such that, for each $k \geqslant 0$, we have $xy^kz \in L(A)$

Example

Let $|y| = m \geqslant 1$

$$w = \overbrace{\underbrace{111\cdots}_{x}\ \underbrace{\cdots 1}_{\substack{y \\ |y|=m\geqslant 1}}\ \underbrace{1111\cdots 11}_{z}}^{p}$$

Choose $k = p - m$, so that $xy^{p-m}z \in L_{pr}$ and then $|xy^{p-m}z|$ is a prime number

Example

We can write $|xy^{p-m}z| = |xz| + (p-m)|y| = p - m + (p-m)m = (1+m)(p-m)$

Let us verify that none of the two factors is a 1 :

- $y \neq \epsilon$, thus $1 + m > 1$
- $m = |y| \leqslant |xy| \leqslant n$, $p \geqslant n + 2$, thus $p - m \geqslant n + 2 - m \geqslant n + 2 - n = 2$

We have derived a **contradiction** $\square$

$ww^R = xyz \quad \begin{cases} |xy| \leq n \\ y \neq \varepsilon \end{cases}$

$|ww^R| = |x| + k|y| + |z| = 2|w|$

- $(|x| + |z|)$ even and $|y|$ not even:

  $\forall k$ not even $\rightarrow xy^kz \notin L$

- $(|x| + |z|)$ not even:

  $\forall k$ even $\rightarrow xy^kz \notin L$

- $(|x| + |z|)$ even and $|y|$ even:

  if $x \neq w$, with $k=0$ $xy^kz \notin L$