CS 4510: Automata and Complexity 09/13/2022

Lecture 7: Nonregular Languages

Lecturer: Zvi Galil *Author: Austin Peng*

Contents

1 Example: Nonregular Language

Example 1. Consider the language $L = \{O^n 1^n \mid n \geq 0\}$.

If we attempt to find a DFA that recognizes L, we find that the machine needs to remember how many 0s have been read so far. Because the number of 0s is not limited, the machine will have to track an unlimited number of possibilities (but this can't be done with finite states). Therefore, L is not a regular language.

Example 2. Consider the language $A = \{w \mid w \text{ has an equal number of 0s and 1s}\}$ and $B = \{w \mid w \text{ has an equal number of occurrences of 01 and 10 as substrings}\}$.
A is not regular but B is. Below, we will use the pumping lemma to show how to prove that certain languages are not regular.

2 Pumping Lemma For Regular Languages

Theorem 1. If A is a regular language, then there is a number p (pumping length) where if s is any string in A of length $\geq p$, then s may be divided into 3 pieces, $s = xyz$, satisfying the following conditions:

- for each $i \geq 0, xy^iz \in A$
- $|y| > 0$
- $|xy| \leq p$

Proof. If A is regular then there exists a DFA M accepting A.

Let $M = (Q, \Sigma, \delta, q_1, F)$ be a DFA recognizing A and p be the number of states of M.

Let $s = s_1s_2...s_n$ be a string in A of length n, where $n \geq p$. Let $r_1, ..., r_{n+1}$ be the sequence of states that M enters while processing s, so $r_{i+1} = \delta(r_i, s_i)$ for $1 \leq i \leq n$. This sequence has length $n + 1$, which is at least $p + 1$. Among the first $p + 1$ elements in the sequence, 2 must be the same state, by the pigeonhole principle. Call the first of these r_j and the second r_l. Because r_l occurs among the first $p+1$ places in a seuqnce starting at r_1, we have $l \leq p+1$.

Now let $x = s_1 \cdots s_j - 1, y = s_j \cdots s_{l-1}, z = s_l \cdots s_n$. As x takes M from r_1 to r_j, y takes M from r_j to r_j, and z take sM from r_j to r_{n+1}, which is an accepting state so M must accept xy^iz for $i \geq 0$. We know that $j \neq l$, so $|y| > 0$, and $l \leq p + 1$, so $|xy| \leq p$, and have satisfied all conditions of the pumping lemma.

$\square$

2.1 Examples: Using Pumping Lemma To Prove Nonregularity

Example 3. Consider the language B be the language $\{0^n 1^n \mid n \geq 0\}$.

Proof. Assume that B is regular.

Let p be the pumping length given by the pumping lemma. Choose s to be a string $0^p 1^p$. Because s is a member of B and s has length more than p, the pumping lemma guarantees that s can be split into 3 pieces, $s = xyz$, where for any $i \geq 0$ the string xy^iz is in B. Let us consider 3 cases to show that this result is impossible:

1. The string y consists only of 0s. For example, $xyyz$ (when $i = 2$) has more 0s than 1s, so it is in the language B. This breaks condition 1 of the pumping lemma.
2. The string y consists only of 1s. This also breaks condition 1.
3. The string y consists of both 0s and 1s. The string $xyyz$ may have the same number of 0s and 1s, but they are out of order. So the string is not in B and breaks condition 1 of the pumping lemma.

$\square$

Example 4. Let $C = \{w \mid w \text{has an equal number of 0s and 1s}\}$. Prove that C is not regular.

Proof. Assume that C is regular.

Let p be the pumping length given by the pumping lemma. Choose s to be the string $0^p 1^p$. Because s is a member of C and s has length more than p, the pumping lemma guarantees that s can be split into 3 pieces, $s = xyz$, where for any $i \geq 0$ the string xy^iz is in C.

Let x and z be the empty string and y be the string $0^p 1^p$. Then xy^iz always has an equal number of 0s and 1s and is in C.

BUT THIS IS WRONG! CONDITION 3 OF THE PUMPING LEMMA COMES INTO PLAY.

By pumping s, it must be divided so that $|xy| \leq p$. That means the string we selected $s = 0^p 1^p$ cannot be pumped. If $|xy| \leq p$, then y must consist only of 0s, so $xyyz \notin C$.

s cannot be pumped and therefore C is not regular. $\square$

Example 5. Consider the language $D = \{ww \mid w \in \{0,1\}^*\}$. Prove that D is not regular.

Proof. Assume that F is regular.

Let s be the string 0^p10^p1. Because s is a member of F and s has length more than p, the pupming lemma guarantees that s can be split into 3 pieces $s = xyz$ satisfying the pumping lemma.

Condition 3 of the pumping lemma is important again. We cannot let x and z be of length 0 because of condition 3. This means that y must consist of only 0s, so $xyyz \notin D$. $\square$

Example 6. Consider the language $\tilde{L_3} = \{w_1w_2 \mid w_1 \neq w_2; w_1, w_2 \in \Sigma^*\}$.

Then $L_3 = (\Sigma\Sigma)^* - \tilde{L_3} = (\Sigma\Sigma)^* \cap \tilde{L_3}$

2.1.1 Example: Pumping Lemma On Unary Languages

Example 7. Consider the language $F = \{1^{n^2} \mid n \geq 0\}$. F essentially contain all strings of 1s whose lengths is a perfect square. We will show that F is not regular.

Proof. By contradiction:

Assume that F is regular. Let p be the pumping length given by the pumping lemma. Let s be the string 1^{p^2}. Because s is a member of F and s has length at least p, the pumping lemma guarantees that s can be split into $s = xyz$, where for any $i \geq 0$, the string xy^iz is in D.

Consider the 2 strings xyz and xy^2z. These strings differ from each other by a single repetition of y and their lengths differ by the length of y. By condition 3 of the pumping lemma, $|xy| \leq p$ and thus $|y| \leq p$. We have $|xyz| = p^2$ and so $|xy^2z| \leq p^2 + p$. But $p^2 + p < p^2 + 2p + 1 = (p+1)^2$. Condition 2 implies that y is an not the empty string so $|xy^2z| > p^2$. The length of xy^2z lies strictly between the consecutive perfect squares p^2 and $(p+1)^2$. So the length cannot be a perfect square, see that $xy^2z \notin F$, and conclude that F is not regular.

$\square$