Till now we learned that for all the regular languages, how to construct an automaton.

But there are a "large class" of languages for which it is not possible to construct an automaton.

The question is "How do we know that it is not possible to design an automaton for a given language"??

In order to answer the above question, we need to understand a very important property of a regular language. This property is described by the following lemma called the Pumping lemma

## Pumping lemma

For every regular language L, there exists a positive integer P, (called the pumping length) such that for every string $w \in L$ with

$|w| \geq p$, there exists strings $x, y, z \in \Sigma^*$ such that $w$ can be written as $w = xyz$, where

(1) $|y| > 0$

(2) $|xy| \leq p$

(3) for every $i \geq 0$, $xy^iz \in L$.

**Proof** Let $L$ be a regular language

Then there exists a DFA $M = \langle Q, \Sigma, \delta, q_0, F \rangle$ such that $L(M) = L$.

Let the number of states in $Q$ is $p$.

Let $w$ be a string of length $n$ where $n \geq p$.

Let $w = a_1 a_2 a_3 \cdots a_n$ where $a_i \in \Sigma$

Define the substring $w_i = a_1 a_2 \cdots a_i$

i.e,
$$w_1 = a_1$$
$$w_2 = a_1 a_2$$
$$w_3 = a_1 a_2 a_3$$
$$\vdots$$
$$w_n = a_1 a_2 \cdots a_n = w$$

Let $\hat{\delta}(q_0, w_1) = r_1$

$\hat{\delta}(q_0, w_2) = r_2$

$\vdots$

$\hat{\delta}(q_0, w_p) = r_p$

and define $q_0 = r_0$

Here we have $p+1$ states $r_0, r_1, \ldots, r_p$.

Since each $r_i \in Q$ and $|Q| = p$

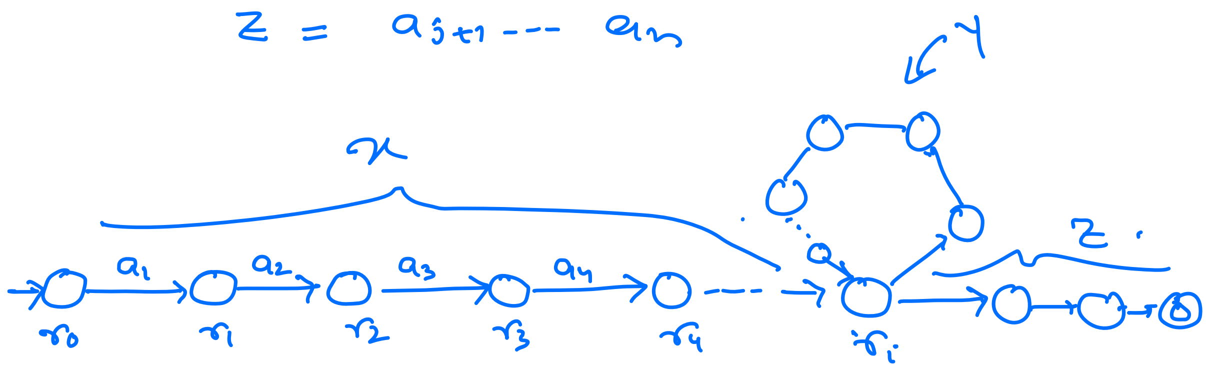
therefore there exists some $i, j$, $i < j$

such that $r_i = r_j = q$ (say).

take $x = a_1 a_2 \cdots a_i$

$y = a_{i+1} a_{i+2} \cdots a_j$

$z = a_{j+1} \cdots a_n$



Claim: For $i \geq 0$ $\hat{\delta}(q_0, xy^i) = \hat{\delta}(q_0, x)$

Proof of the claim (using induction)

for $i=0$, trivially true.

for $i=1$, $\hat{\delta}(q_0, xy) = r_j = r_i = \hat{\delta}(q_0, x)$

Suppose the claim is true for $i = l$

We have

$$\hat{\delta}(q_0, xy^{l+1}) = \hat{\delta}(q_0, xy^l y)$$

$$= \hat{\delta}(\hat{\delta}(q_0, xy^l), y)$$

$$= \hat{\delta}(\hat{\delta}(q_0, x), y)$$

$$= \hat{\delta}(q_0, xy)$$

$$= \hat{\delta}(q_0, x)$$

Now we are ready to complete the proof of the pumping lemma.

Since $w \in L$, $\hat{\delta}(q_0, w) \in F$

Now $\hat{\delta}(q_0, xy^iz)$

$= \hat{\delta}(\hat{\delta}(q_0, xy^i), z)$

$= \hat{\delta}(\hat{\delta}(q_0, x), z)$

$= \hat{\delta}(q_0, xz)$

$\hat{\delta}((\hat{\delta}(q_0, xy)), z)$

$= \hat{\delta}(q_0, xyz) \in F$

This shows that $xy^iz \in L$

---

Pumping lemma shows a property ~~for~~ regular languages.

Does it says anything about non - regular languages ?!

Ans - NO

Then How is it helpful in our context??

**Q.** What if a language doest satisfy Pumping lemma? Can it be regular??

The answer is No.

| The method to show a language is not regular |
| --- |

→ Assume it is regular. so Pumping lemma holds

→ Show that for some $w \in L$, for any choice of $x, y, z$, $xy^iz \notin L$

**Example 1** Show that the language $\{0^n 1^n \mid n \geq 0\}$ is not regular.

**Sol^n**

Let the language is regular.

This implies that Pumping lemma holds.
Let $p$ be the pumping length.

Consider the string $w = 0^p 1^p$.

As per the statement of Pumping lemma,

$\exists\, x, y, z \in (0+1)^*$ s.t $w = xyz$ with $|xy| \leq p$ and $|y| \geq 1$ and $xy^iz \in L$ for $i \geq 0$.

Now, since $|xy| \leq p$, for any choice of $x, y$, the string $xy$ cannot contain any 1.

i.e $xy = 0^m$ for some $m \leq n$.

Suppose that $x = 0^{k_1}$, $y = 0^{k_2}$ for some $k_1, k_2$, $k_2 \neq 0$

Then $xy^2z = 0^{n+k_2} 1^n$

since $k_2 \neq 0$, $xy^2z \notin L$.

This is a contradiction.

Therefore $L$ is not regular.

## Example 2

Show that the language $\{ w \in \{a+b\}^* \mid |w|_a = |w|_b \}$ is not regular.

## Example 3

Show that the language $L_p = \{ 0^i \mid i \text{ is a prime} \}$ is not regular.

## Example 4

Show that the language $L_p = \{ 1^{n^2} \mid n \geq 0 \}$ is not regular.

## Example 5

Show that the language $L_p = \{ xx \mid x \in \{0,1\}^* \}$ is not regular.

Solutions of the above problems are discussed in the class. They also can be found in the book.

## Important note

Remember that for a carefully chosen string $w$, your proof must show that there exist a $i \geq 0$ s.t $xy^iz \notin L$ for all possible choices of $x, y, z$