A closer look at regular languages:

reminder:

reg. language ⇔ finite automaton

⇕

reg. expression ⇔ finite automaton

- now let's look at the fact that reg. languages are recognized by finite automata.

- finite automaton = finite amount of states.

yet, many regular languages contain an infinite amount of strings and strings that can grow infinitely large.

how is that possible?

↓

it is made possible thanks to loops.

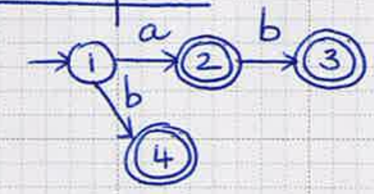
what this means:

(*) if a FA with p states recognizes a string of length $\geq p$, then the FA has a loop.

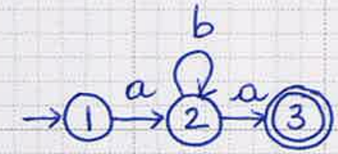
---

* if it didn't have a loop (i.e., a state that is repeated), the max. length of accepted strings would be: $p-1$.

examples:



this FA recognizes $\{a, b, ab\}$.
it has 4 states.
no loop: finite amount of strings, each of finite (and < 4) length.



this FA recognizes ab^*a.
it has 3 states → max length without repeats = 2.
yet aba has length 3 and is accepted → there is a loop (a state - state 2 - is repeated).

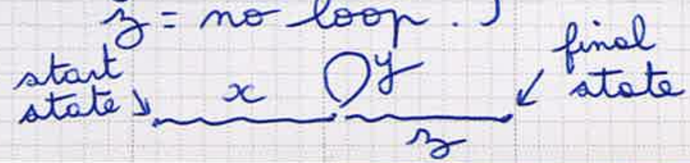
---

now, let's go back to (*):

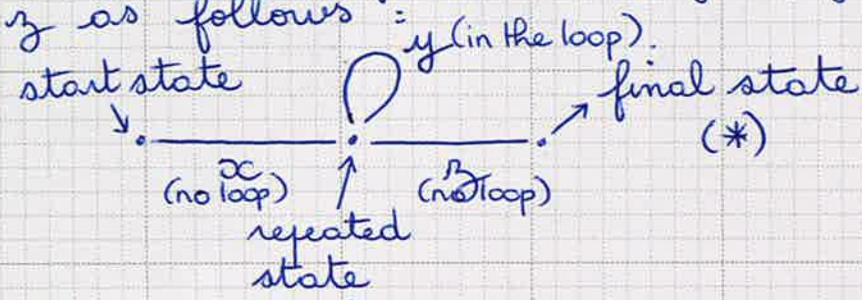
given a DFA with p states, if it recognizes a string w of length $\geq p$ then w goes through a loop in the DFA:

w can be decomposed as 3 parts:

$$\left.\begin{array}{l} x = \text{no loop} \\ y = \text{the loop} \\ z = \text{no loop}. \end{array}\right\} w = xyz$$



since w can be viewed as the concatenation of 3 substrings x, y, z as follows :



(*)

it means that :

length of $y > 0$:

otherwise, since there is no loop in x or z, the length of w would be $< p$.

very important :

from the above diagram (*), we can also conclude that :

xz is recognized by FA

$xyz = w$ is recognized

$xyyz$ is recognized

$xyyyz = xy^3z$ is recognized

xy^4z is recognized .

$\vdots$

xy^iz, for any $i \geq 0$, is recognized

now you are ready for the following property, which we have pretty much dissected in the previous pages. It is called the :

— pumping lemma —

if A is a regular language, then : there exists a length p above which all strings s can be divided into 3 pieces :

$$s = x\,y\,z$$

where :

$|y| > 0$ (i.e., $y \neq \varepsilon$)

$|xy| < p$

for all $i \geq 0$, $xy^iz \in A$.

* p is called the pumping length and a good value for p is the number of states of an FA that recognizes A.