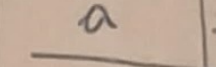
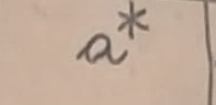
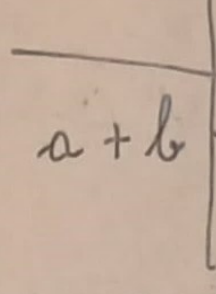
04.2024

# Problems with simplistic parser in regex to ε-NFA

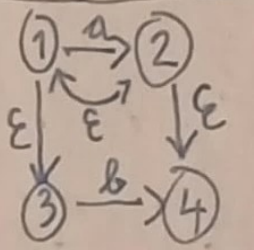
by Rescariu Matei

initial ideas: (wrong)

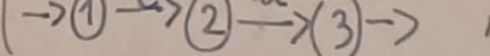
| regex | ε-NFA |
|---|---|
| a |  |
| a* |  |
| a+b |  |

pros: simple, easily trackable (no changing previous decisions), parses from left to right.

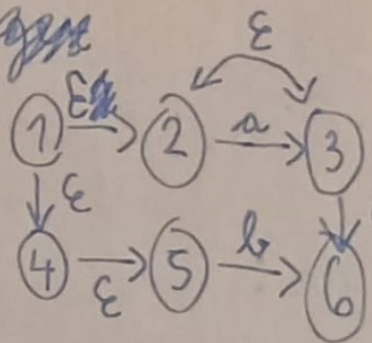
PROBLEMS:

| | regex | old-solution | problem ~~solution~~ |
|---|---|---|---|
| (1) 1st noticed problem | (a*)+b |  | accepts extra words. Example: "a b" |

It can reach "b" after backtracking on "a" which isn't intended in the regex. Therefore, I came with a solution:

| regex | ε-NFA |
|---|---|
| a |  |

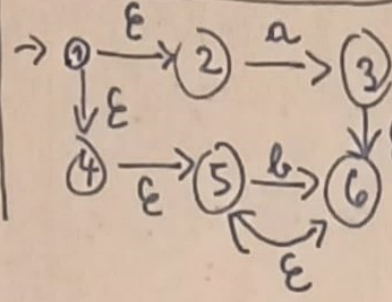
, meaning the graph for the (a*)+b is:



Now it works on this input. Still, it fails on the next (on pg. 2).

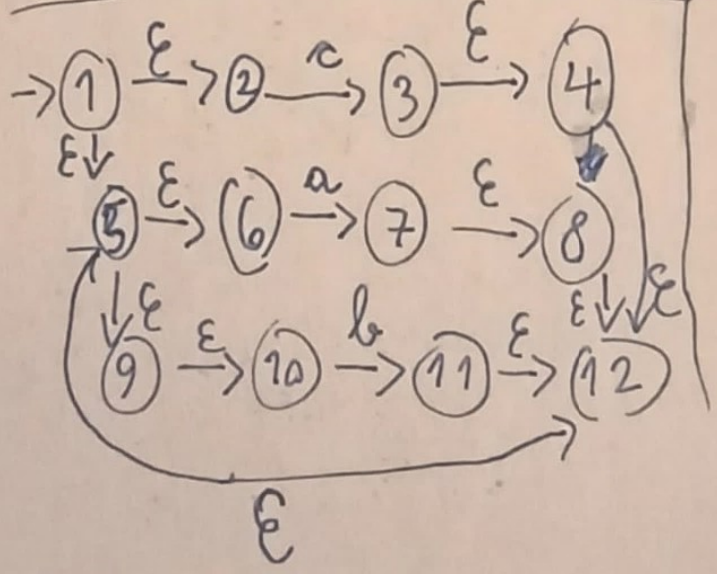
04.2024

# Evoluția parsării simplificate a mea
(Pescarin Matei)

| | regex | to ε-NFA | problem |
|---|---|---|---|
| (2) | a+b* |  | Similarly, it accepts "a b". |

(Ugly) solution: change "a" to ① $\xrightarrow{\varepsilon}$ ② $\xrightarrow{a}$ ③ $\xrightarrow{\varepsilon}$ ④.

Now, the problems seem all fixed. NOPE!

| (3) | regex | to ε-NFA | problem: |
|---|---|---|---|
| | c+(a+b)* |  | accepts "ca" |

solution: choose a different method.

---

Moral: be very careful when taking a shortcut/simplified solution.