

Regexes $\rightarrow$ NFAs

Recursive, based on how the regex is constructed.

Base cases:

regex	NFA
$\emptyset$	$\rightarrow \bigcirc$
ϵ ("")	$\rightarrow \bigcirc \bigcirc$
a (ascii)	$\rightarrow \bigcirc \xrightarrow{a} \bigcirc$

$r | s$
 rs
 r^*



Fact: This transformation is efficient (linear time)



Simulating an NFA

Represent the NFA A

by a transition table δ

For every state q and

char a ,

$\delta(q, a)$ is the set of all possible states reachable from q by reading the character a .



$\delta(q, a) = \{t, u, v, y, x, z\}$

compute efficiently

$\delta(q, a)$ for all states q

and all chars a .

Simulating the NFA A :

state := the set of states

reachable from the start state

after reading the input.

// state := the set of all

states reachable from state_i

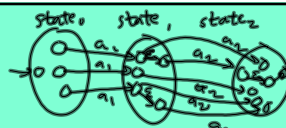
compute state_{i+1} from state_i

for each $i \geq 0$:

Suppose input is $a_0 a_1 \dots a_n$

each a_i is an ascii char.

state_{i+1} := $\bigcup_{q \in \text{state}_i} \delta(q, a_{i+1})$



state_n is the set of all possible states readable by reading the entire input.

Accept iff at least one of the states in state_n is accepting.

$$\bigcup_{q \in \text{state}_n} \delta(q, a_1 \dots a_n)$$

NFA $\rightarrow$ DFA

A DFA is an NFA where, for every state q and every char a , there is exactly one edge leaving state q with label a . (+ no ϵ -moves either)

(much easier to simulate (real time!))

NFA $A \xrightarrow{\text{equiv}} \text{DFA } B$

States of B are sets of states of A .

Transition function table Δ for B is as follows:

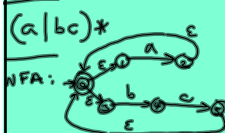
$$\Delta\left(\underbrace{\text{set of states of } A}_S, \underbrace{\text{ascii char}}_a\right) = \underbrace{\bigcup_{q \in S} \delta(q, a)}_{\text{set of states of } A = \text{a state of } B}$$

Start state of B

= state₀ = set of states of A reachable from start state of A following 0 or more ϵ -moves only.

State S of B is a final (accepting) state iff it contains at least one accepting state of A .

Usually, not all possible subsets are used (unreachable from start state of B). Do not need to include these.



Δ table for B , labelling start & final states

state	a	b	c
* 013	2013	4	$\emptyset$
* 2013	2013	4	$\emptyset$
4	$\emptyset$	$\emptyset$	5013
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
* 5013	2013	4	$\emptyset$

