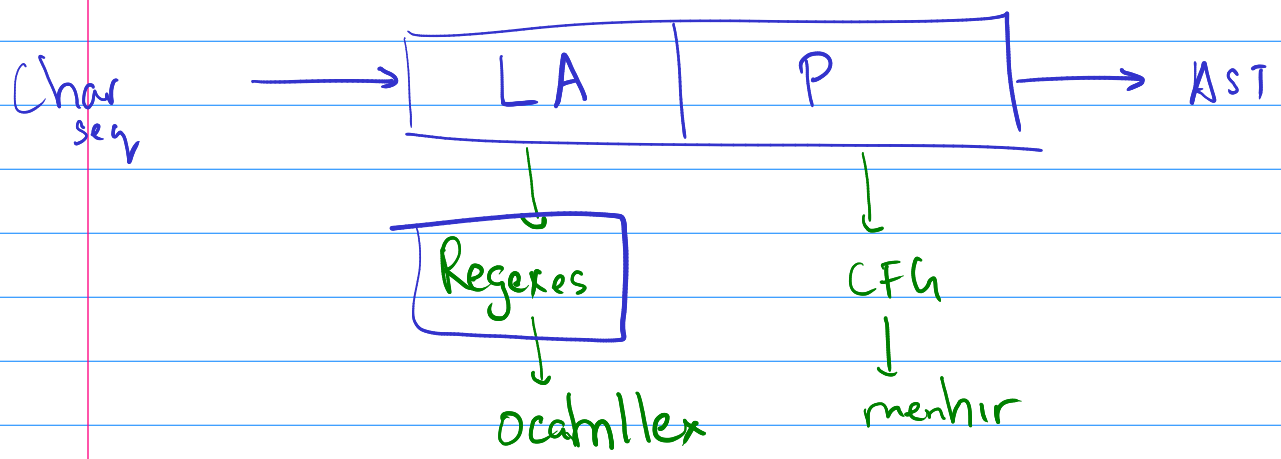


$$r ::= \bot \mid \varepsilon \mid a \mid r_1 + r_2 \mid r_1\, r_2 \mid r^*$$

Base cases — Disjunction, Concat., Kleene-*.

Alphabet $\Sigma = \{ a \; b \; c \; \cdots \; z \}$ ← Finite alphabets

- $\Sigma = a + b + c + \cdots + z$ ← Special regex which recognizes any 1-character string

Σ^* — accepts any string that can be built out of the alphabet.

$a\Sigma^* \qquad \Sigma^* b \qquad \Sigma^* a \Sigma^* \qquad \Sigma^* ab \Sigma^*$

$$\Sigma^* (a + e + i + o + u) \Sigma^*$$

① English language desc $\rightleftarrows$ Regex / CFG

Language (r) = set of strings that r accepts

$$[\![r]\!]$$

② Given regex r string w, does $w \in [\![r]\!]$?

Is there an algorithm for this problem?

ASIDE: $r^* = 0$ or more repetitions of r

$r^+ = 1$ or more repetitions of r

$= r \cdot r^*$

If I had defined regexes as

$r ::= \perp \mid \varepsilon \mid a \mid r_1 + r_2 \mid r_1 r_2 \mid r^+$ || $r^* = \varepsilon + r^+$

$r? = r + \varepsilon$

$[\![r]\!]$ = set of set of strings accepted by r.

Problem: Given r (regex), w, does $w \in [\![r]\!]$?

w match the regex?

Problem: Given r, w, does w match r?

Abhishek's suggestion

M's question: How many strings do we produce?

Produce all strings generated by r

Check if w occurs in this list.

M's response to M's own question

The only strings that we need to produce are those of length $\leq |w|$.

$$\frac{1-|\Sigma|^{|w|+1}}{1-|\Sigma|} = |\Sigma|^0 + |\Sigma|^1 + \cdots + |\Sigma|^{|w|}$$

Sathwika's suggestion: Check if $r + w = r$.

$$[\![r + w]\!] = [\![r]\!]$$

$$r ::= \bot \mid \varepsilon \mid a \mid r_1 + r_2 \mid r_1 r_2 \mid r^*$$

matches r w =

If $r = \bot$ then false

else if $r = \varepsilon$ then $w = \varepsilon$?

else if $r = a$ then $w = a$?

else if $r = r_1 + r_2$ then matches r_1 w or matches r_2 w

else if $r = r_1 \cdot r_2$ then ...

w

pre	suff
$\leq \lvert w \rvert$	$\leq \lvert w \rvert$



Does this substring match this subregex?

$$\Sigma^* a^+ \Sigma^* \longrightarrow \Sigma, \Sigma^*, a, a^+, \Sigma^* a^+, \Sigma^* a^+ \Sigma^*$$

Memoization table has $O(|r||w|^2)$ entries

Filling in each entry takes $O(|w|)$ time

Simple dynamic programming gives a $O(|r||w|^3)$ time algorithm for regex matching

$$(10\,000)^3 = (10^4)^3 = 10^{12}$$

Example: Does a string match Hi?

```
#!/usr/bin/env python

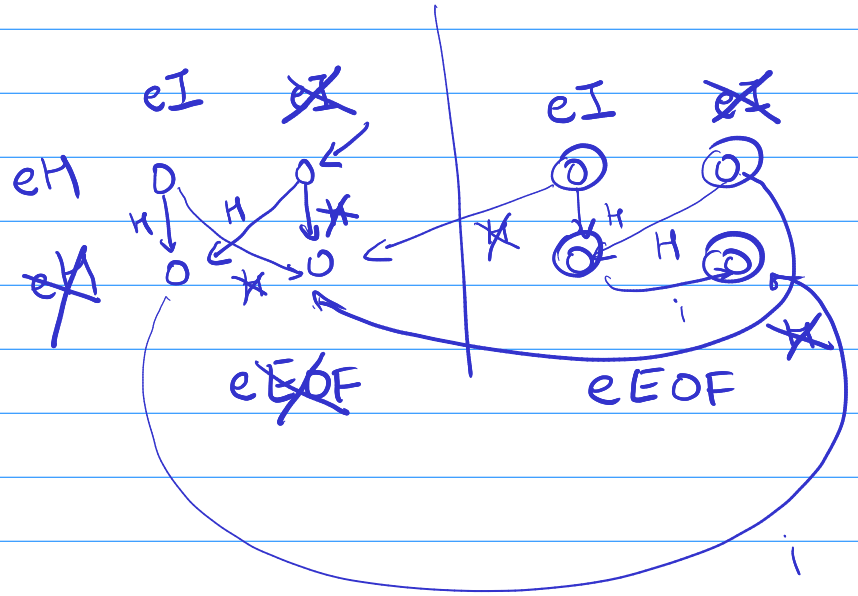
# r = Hi

def matchesHi(w: str) -> bool:
    expectingH = True
    expectingI = False
    expectingEOF = False

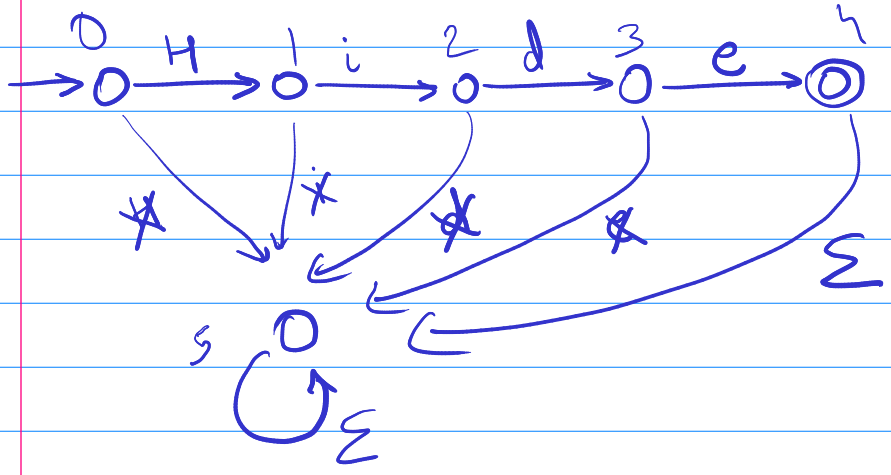
    for c in w:
        if expectingH and c == 'H':
            expectingH = False
            expectingI = True
        elif expectingI and c == 'i':
            expectingI = False
            expectingEOF = True
        else:
            expectingH = False
            expectingI = False
            expectingEOF = False

    return expectingEOF

for w in ['', 'H', 'h', 'Hi', 'hi', 'Hii', 'hHi']:
    print(f'{w}: {matchesHi(w)}')
```



Example: Does a string match Hide?



Example: $(Help)^*$

Example: $H_1 + Help$

Example: $(H_1 + Help)^*$

Example: $H_1 + Help$

