CS & IT ENGINEERING

Theory of Computation

Regular Expression-1

Lecture No. 2

By- DEVA Sir

Introduction to TOC

TOPICS TO BE COVERED

01 Regular Expression

02 Operators

03 Basic Regular Expressions

04 Simplification of Reg Exps

05 Writing Regular Exps

Language

↳ Set of Strings

$\Sigma = \{a, b\}$

Languages over Σ:

$\{\ \}$ $\boxed{\{a, b\}}$ $\{\varepsilon, a, b\}$

$\{\varepsilon\}$ $\{a, \varepsilon\}$

$\{a\}$

$\{b\}$

⋮ ⋮

Language Vs Alphabet

Language ↳ set

Alphabet ↳ set

→ Alphabet is language.

→ Language need not be Alphabet

$$L = \{\boxed{a}, b\}$$

a → symbol

a → string

a → one length string

$\Sigma = \{a, b\}$

Symbol Vs String

aaa

aab

⋮

aa → not symbol

aa → String

ab, ba, bb, ...

ε → String

ε → not symbol in Σ

a → Symbol

a → String

b → symbol

b → String

Every symbol is string.

String need not be symbol

$\Sigma = \{a, b\}$

No. of symbols in $\Sigma \Rightarrow$ finite

$\varepsilon, a, b, aa, ab, ba, bb, \ldots$

No. of strings over $\Sigma \Rightarrow$ Infinite

$\{\}, \{\varepsilon\}, \{a\}, \{b\}, \{aa\}, \ldots$

$\ldots$

No. of languages over $\Sigma \Rightarrow$ Infinite

$$\Sigma = \{\varepsilon, a\}$$

symbol

one length string

$|\varepsilon| = 1$

$|a| = 1$

ε is not empty string

you have take some other notation

Assume λ is empty string

$|\lambda| = 0$

$\Sigma = \{ \varepsilon, \lambda, a, b \}$

symbol
one length

empty string $=$ $

Alphabet

$\Downarrow$

Strings

$\Downarrow$

Languages

Operations on Strings

concatenation	$w_1 = ab,\ w_2 = aaa$ $\left.\right\}$ $w_1 \cdot w_2 = abaaa$, $w_2 \cdot w_1 = aaaab$
Reversal	$w = abc \Rightarrow w^{Rev} = cba$
prefix(abc)	ϵ, a, ab, abc — Beginning Sequence
suffix (abc)	ϵ, c, bc, abc — Ending Sequence
substring (abc)	ϵ, a, b, c, ab, bc, abc — Part of string
subsequence (abc)	ϵ, a, b, c, ab, bc, ac, abc — any subsequence
Length	$w = abc \Rightarrow \lvert w \rvert = 3$

Unary

Binary

cb

is not substring

ca

is not subsequence

$w = abcd$

$\text{prefix}(w) = \{u \mid uv = w\}$

ε
a
ab
abc
$abcd$

5 prefixes

$\text{Suffix}(w) = \{v \mid uv = w\}$

ε
d
cd
bcd
$abcd$

5 suffixes

$\text{Substring} = \{y \mid xyz = w\}$

ε
a
b
c
d
ab
bc
cd
abc
bcd
$abcd$

Subsequence

2 choices, 2, 2, 2 $\Rightarrow 2^4 = 16$

$\varepsilon\ \varepsilon\ \varepsilon\ \varepsilon = \varepsilon$
$a\ \varepsilon\ \varepsilon\ \varepsilon = a$
$\varepsilon\ b\ \varepsilon\ \varepsilon = b$
$\varepsilon\ \varepsilon\ c\ \varepsilon = c$

ε	abc
a	abd
b	acd
c	bcd
d	$abcd$
ab	
ac	
ad	
bc	
bd	
cd	

$w = abcd$

$\text{Prefixes}(w) = \{u \mid uv = w\}$

$= \{\varepsilon, a, ab, abc, abcd\}$

$w = abcd$

$\boxed{u \; v = w}$

u	v		w
ε	$abcd$	$=$	$abcd$
a	bcd	$=$	"
ab	cd	$=$	"
abc	d	$=$	"
$abcd$	ε	$=$	"

n Length string

- No. of prefixes = $n+1$
- No. of suffixes = $n+1$
- *** No. of Substrings = ($\boxed{\text{min}}$, $\boxed{\text{max}}$)
  - min (when all symbols are same) $\Downarrow$ $n+1$
  - max (when all symbols in string are distinct) $\Downarrow$ $\Sigma n+1$
- No. of subsequences = (min , max)
  - min $\Downarrow$ $n+1$
  - max $\Downarrow$ 2^n

$w = \underbrace{aaaa}_{4}$

Substrings $\left\{ \begin{array}{l} \varepsilon \\ a \\ aa \\ aaa \\ aaaa \end{array} \right.$

$= 5$

$w = \underbrace{abcd}_{4}$

n length

$\varepsilon \rightarrow 1$

$\left.\begin{array}{l} a \\ b \\ c \\ d \end{array}\right\} 4$

$\left.\begin{array}{l} ab \\ bc \\ cd \end{array}\right\} 3$

$\left.\begin{array}{l} abc \\ bcd \end{array}\right\} 2$

$abcd \} 1$

$1+2+3+4$

$1+2+3+4+\dots+n$

$\underbrace{1+2+3+4+\dots}_{\sum n = \frac{n(n+1)}{2}}$

$(1+2+3+4)+1$

$\frac{n(n+1)}{2}+1$

Operations on Strings

concatenation :

$$\left.\begin{array}{l}|w_1| = n_1 \\ |w_2| = n_2\end{array}\right\} \Rightarrow |w_1 . w_2| = \overset{2+3}{n_1 + n_2}$$

$$\left.\begin{array}{l}w_1 = ab \\ w_2 = acd\end{array}\right\} \Rightarrow w_1 w_2 = \underbrace{ab.acd}_{5}$$

Prefix

Suffix

Substring

Subsequence

Length of string

Operations on Strings

Concatenation of 2 strings

$w_1 = abc$
$w_2 = ef$

$w_1 . w_2$

w_1 followed by w_2

$$\boxed{w_1 w_2 = abc.ef}$$

Operations on Sets

(Language)

- Union of 2 sets
- Intersection of 2 sets
- Complement of a set
- Concatenation of 2 sets
- Reversal of a set

⋮

Operations on Sets

$L_1 = \{ab, baa\}$ $\quad L_2 = \{\varepsilon, a\}$

$L_1 \cup L_2 = \{w \mid w \in L_1 \text{ or } w \in L_2\}$

$= \{ab, baa, \varepsilon, a\}$

What is TOC?

Problems

Computable

Not Computable

How to represent? How much harder?

Automata, Computability, and Complexity

(machines)

* I. II III

Computable problem

TOC ⇒ 3 branches

Theory of Computation

Automata Theory

Formal Languages

FLAT

Chomsky Hierarchy

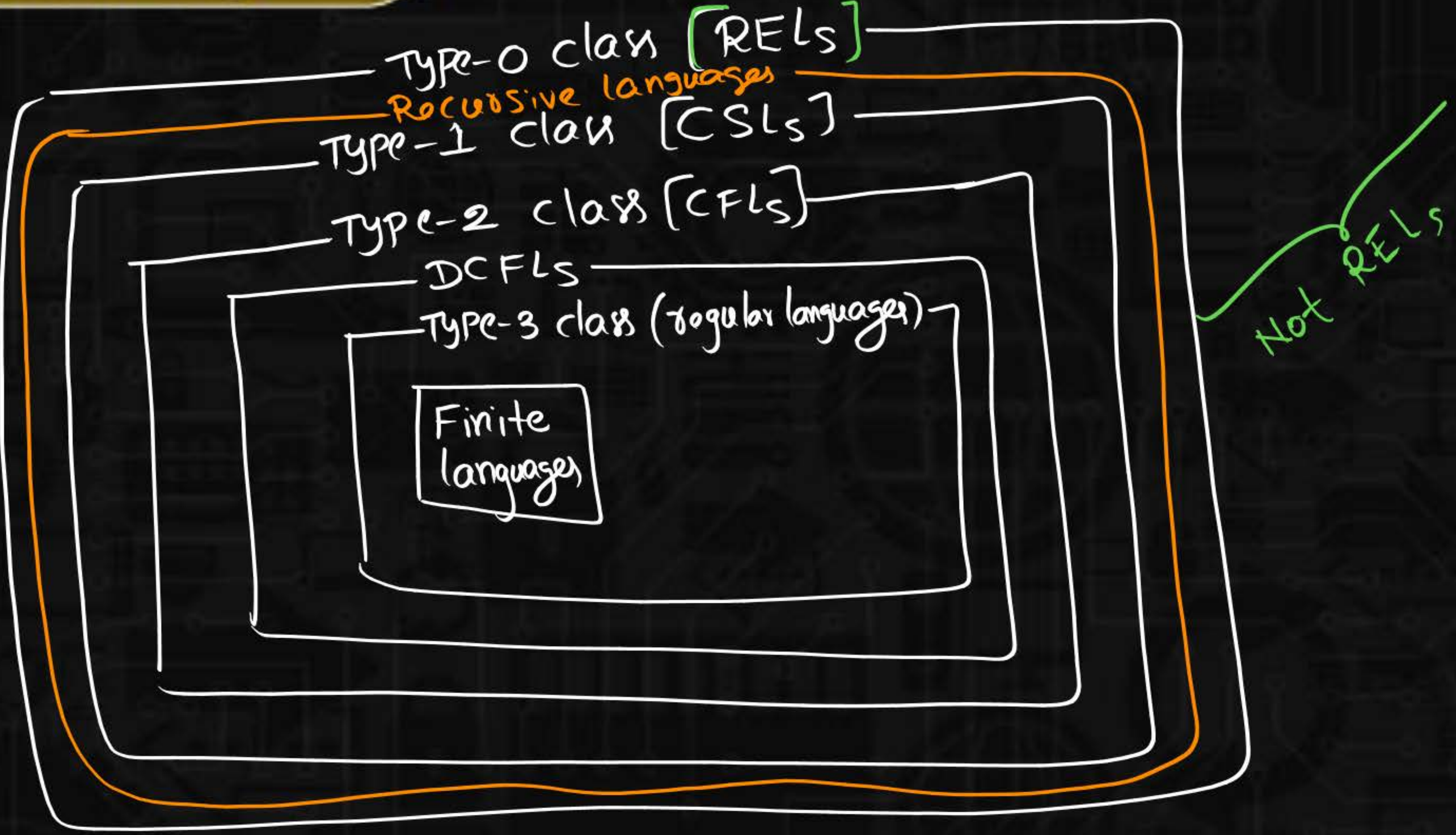
⇒ 4 classes [T-3, T-2, T-1, T-0]

REL
↓
Recursively Enumerable

CSL
↓
Context Sensitive language

DCFL
↓
Deterministic CFL

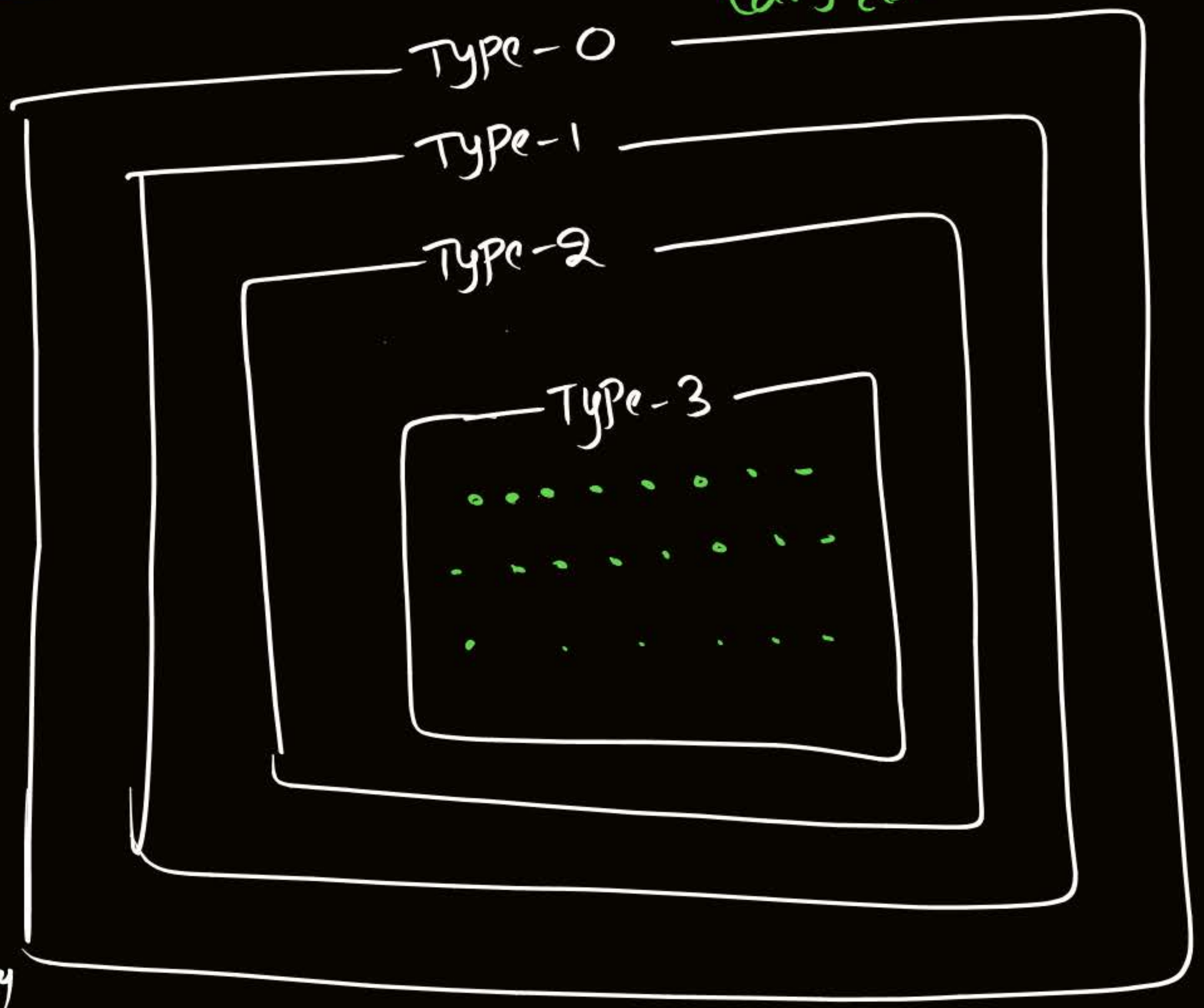
CFL
↓
Context Free Language



Class: $\boxed{\text{Type-3} \subset \text{Type-2} \subset \text{Type-1} \subset \text{Type-0}}$

Type-3 → Smallest class

Type-0 → Largest class



class

↳ Collection of languages

Type-3 class:

↳ Collection of regular languages

Classes	Type-3	Type-2	Type-1	Type-0
Language	Regular	CFL	CSL	REL
Automata (machine)	Finite Automata	PDA (Push down Automata)	LBA (Linear Bound Automata)	TM (Turing machine)
Grammar	Regular Grammar → LLG, RLG	CFG	CSG	UG (Unrestricted)

What is Language?

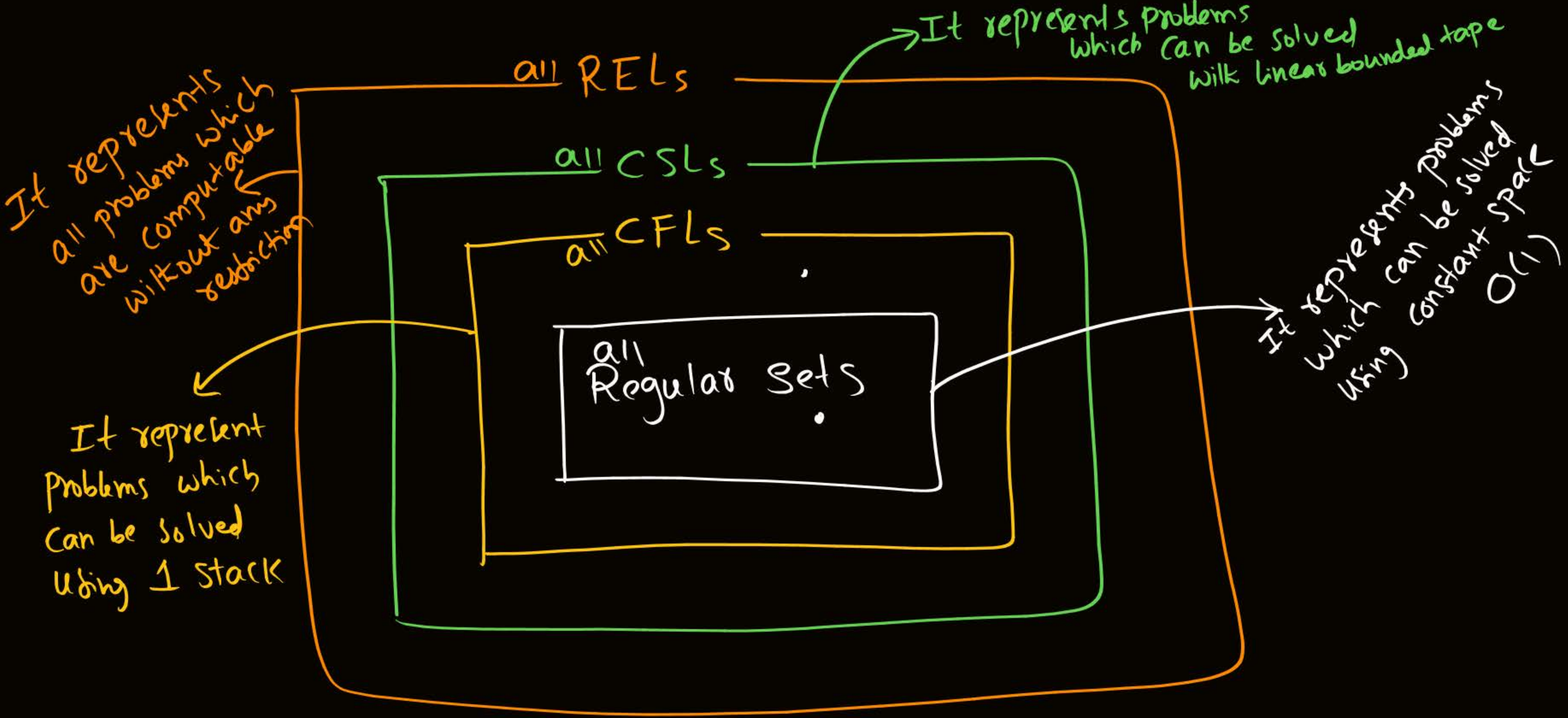
collection of strings

What is Automata?

It is a machine that represents a language

What is Grammar?

It is set of rules that generates a language


It represents all problems which are computable without any restriction
all RELs
It represents problems which can be solved with linear bounded tape
all CSLs
all CFLs
all Regular Sets
It represents problems which can be solved using constant space O(1)
It represent problems which can be solved using 1 stack

→ Every Regular language is CFL
is CSL
is REL

→ Every CFL is CSL
is REL

→ Every CSL is REL

CFL need not be regular

CSL need not be CFL

REL need not be CSL

Q.1 Let $L = \{\varepsilon, a, ab, aba\}$. Then find length of smallest string

ε ⇓ min, a ⇒ smaller, aba ⇓ largest string

A. 0 ✓

B. 1

C. 2

D. 3

$|\varepsilon| = 0$

Q.2 Find correct statement.

$a.b = ab$ $\quad a.\varepsilon = a$

A. $\{\boxed{a}, \boxed{ab}\} \cdot \{\boxed{\varepsilon}, \boxed{b}\} = \{a, ab, abb\}$

$ab.\varepsilon = ab$ $\quad ab.b = abb$

B. $\{a, ab\} \cdot \{\varepsilon, b\} = \{a, \cancel{\boxed{\varepsilon}}, \cancel{\boxed{b}}, ab\}$

C. $\{a, ab\} \cdot \{\varepsilon, b\} = \{a, \cancel{b}, abb, ab\}$

D. None of these

Q.3 Let $L = \{aba, aabab\}$. Then prefix(L) = ?

$= \{\text{pref}(aba), \text{pref}(aabab)\}$

A. $\{\varepsilon, a, ab, aba, aa, aab, aaba, aabab\}$

B. $\{a, ab, aba, aa, aab, aaba, aabab\} \subset \text{prefix}(L)$ — ε not present

C. $\{\varepsilon, a, ba, aba, b, ab, bab, abab, aabab\}$

D. None

40 M easy

100 M

≅ 70 M

1 — 50 Rank

Q.4 Let $\Sigma = \{\varepsilon, a, b\}$. Then find correct statement

symbol

not empty string

1 length

A. $|\varepsilon| = |a| = |b| = 1$

$|\varepsilon a| = 2$

symbol

B. $\cancel{|\varepsilon| = 0}$

C. $\cancel{\underbrace{\varepsilon a} = a}$

not empty string

$\cancel{\varepsilon a = a}$

symbol

D. $\cancel{\varepsilon b = b}$

one length symbol

Q.5 $\Sigma = \{a, b\}$ Then

A. $|\varepsilon| = |a| = |b|$ — ε: zero length; a, b: 1 length — $0 = 1 = 1$

B. $|\varepsilon| = 0$ ✓

C. $\varepsilon a = a$ ✓

D. $\varepsilon b = b$ ✓

Q.6 Let $\Sigma = \{a, b\}$. Then $\Sigma.\Sigma = \Sigma^2 = \; ?$

A. $\{aa, bb\}$

B. $\{a, b\}$

C. $\{\varepsilon, a, b\}$

D. $\{aa, \underline{ab}, \underline{ba}, bb\}$ ✓

$\Sigma.\Sigma = \{\boxed{a}, \boxed{b}\}.\{\boxed{a}, \boxed{b}\}$

b.a b.b

a.a a.b

$= \{aa, \underline{ab}, \underline{ba}, bb\}$

Σ^2 = Set of all 2 length strings

$\Sigma^0 = \{\varepsilon\}$ | $\Sigma^1 = \Sigma = \{a, b\}$ | Σ^2 = set of all 2 length strings

Σ^k = Set of all k length strings

$\Sigma = \{a, b\}$

$$|\Sigma^K| = |\Sigma|^K$$
$$= 2^K$$

$$\boxed{\begin{aligned}\Sigma^{\overset{\rightarrow \text{any}}{*}} &= \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \cdots \\ &= \text{Set of all Strings}\end{aligned}}$$

Q.7

Regular language is

A. CFL

B. CSL

C. REL

D. None

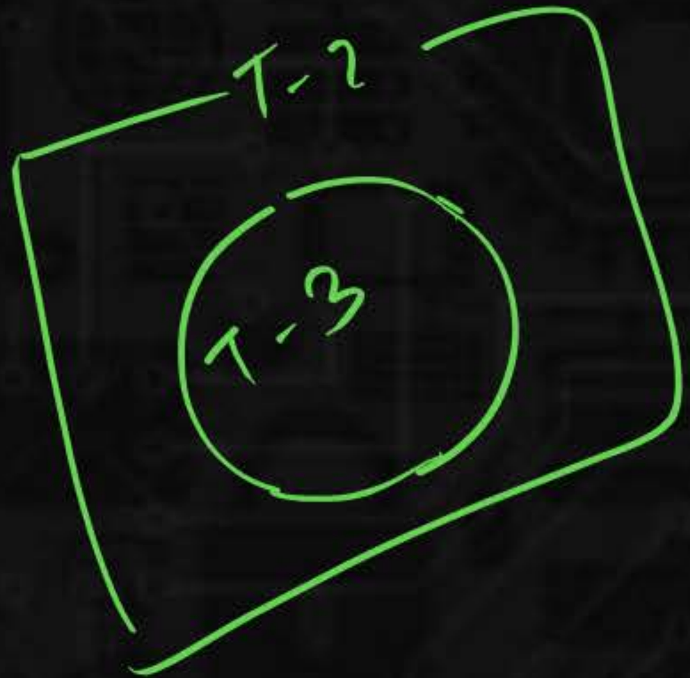
Q.8 $\boxed{\text{Set of regular languages}}$ is same as ______

A. Type-3 class

B. Type-2 class

C. Type-1 class

D. Type-0 class



Q.9 Regular language is $\boxed{\text{equivalent}}$ to

A. Set of all regular languages $= \{ L_1, L_2, L_3, \ldots \}$

B. Belongs to Type-3 class

C. CFL

D. CSL

Regular lang

set of regular

Set of all regulars

→ Regular language

CFL ~~is~~ Regular

$X \equiv Y$

$\boxed{\begin{array}{c} X \Rightarrow Y \\ Y \Rightarrow X \end{array}}$

Regular is CFL

Regular $\not\equiv$ CFL

Q.10 Regular language is belongs to ________

A. ~~Set of all regular languages~~ = Type-3 class

B. ~~Regular language~~

C. ~~Set of all CFLs~~

D. ~~Set of all CSLs~~

$x \in \{a, x, b\}$

$L \in \{A, L, B\}$

$L \in M$

Summary

→ Basics ✓
→ Introduction

⇒ Next: Regular Expression

Subsequences of w

$|w| = n$

$\min = n+1 \longleftrightarrow \max = 2^n$

$w = aaaa$

Subsequences

ε

a

aa

aaa

$aaaa$

$w = abcd$

$\Downarrow$

16 subsequency

2^n


THANK YOU GW
SOLDIERS !