

**IF5110 Teori Komputasi**

## **2. Review Teori Bahasa Formal dan Otomata**

Oleh: Rinaldi Munir

**Program Studi Magister Informatika STEI-ITB**

# Terminologi

- **Alfabet:** himpunan terbatas simbol-simbol

Contoh: alfabet latin , {a, b, c, ..., z}

alfabet Yunanti, { $\alpha$ ,  $\beta$ ,  $\gamma$ , ...,  $\omega$ }

alfabet biner, {0, 1}

- **String:** barisan yang disusun oleh simbol-simbol alfabet.

$$a_1a_2a_3\dots a_n, \quad a_i \in A \quad (A \text{ adalah alfabet})$$

Nama lain untuk string adalah **kalimat** atau ***word***

- Jika  $A$  adalah alfabet, maka  $A^n$  menyatakan himpunan semua string dengan panjang  $n$  yang dibentuk dari himpunan  $A$ .
- $A^*$  adalah himpunan semua rangkaian simbol dari himpunan  $A$  yang terdiri dari 0 simbol (string kosong), satu simbol, dua simbol, dst.

$$A^* = A^0 \cup A^1 \cup A^2 \cup \dots$$

Contoh: Misalkan  $A = \{0, 1\}$ , maka

$$A^0 = \{\epsilon\}$$

$$A^1 = \{0, 1\}$$

$$A^2 = \{11, 01, 10, 11\}$$

$$A^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$$

...

- **Bahasa** (pada alfabet  $A$ ) adalah himpunan bagian dari  $A^*$ .

Contoh: Misalkan  $A = \{a, b, c\}$ , maka berikut ini adalah contoh-contoh bahasa pada alfabet  $A$ :

$$L_1 = \{a, aaa, bc, ac, abc, cab\}$$

$$L_2 = \{aba, aabaa\}$$

$$L_3 = \{\epsilon\}$$

$$L_4 = \{a^i c b^i \mid i \geq 1\}$$

- **Tata bahasa** (*grammar*) adalah aturan yang digunakan untuk membangkitkan atau mengenali kalimat di dalam suatu bahasa.

Contoh tata bahasa Inggris:

**<sentence> → <noun phrase><verb phrase>**

**<sentence> → <noun phrase><verb phrase><noun phrase>**

**<noun phrase> → <adjective><noun phrase>**

**<noun phrase> → <adjective><singular noun>**

**<verb phrase> → <singular verb><adverb>**

**<verb phrase> → <singular phrase>**

**<adjective> → *the***

**<adjective> → *a***

**<adjective> → *little***

**<singular noun> → *boy***

**<singular noun> → *dog***

**<singular verb> → *runs***

**<singular verb> → *bites***

**<adverb> → *quickly***

Contoh pembangkitan kalimat:

**<sentence> → <noun phrase><verb phrase>**

→ **<adjective><noun phrase><verb phrase>**

→ *the* **<noun phrase><verb phrase>**

→ *the* **<adjective><singular noun>< verb phrase>**

→ *the little* **<singular noun>< verb phrase>**

→ *the little boy* **< verb phrase>**

→ *the little boy* **<singular verb><adverb>**

→ *the little boy runs* **<adverb>**

→ *the little boy runs quickly*

- Unsur-unsur tata bahasa:
  1. Himpunan berhingga terminal,  $T$
  2. Himpunan berhingga non terminal,  $N$
  3. Himpunan berhingga aturan produksi,  $P$
  4. Simbol awal,  $S \in N$

Dilambangkan dengan  $G = (T, N, P, S)$

Contoh: tata bahasa  $G = (T, N, P, S)$ , dengan  $T = \{a, b\}$ ,  $N = \{S, A, B\}$ ,  $P = \{S \rightarrow ABa, A \rightarrow BB, B \rightarrow ab, AB \rightarrow b\}$ . Simbol awal adalah  $S$ .

String *abababa* diturunkan sebagai berikut:

$S \Rightarrow ABa$   
 $\Rightarrow Aaba$   
 $\Rightarrow BBaba$   
 $\Rightarrow Bababa$   
 $\Rightarrow abababa$

- **Kelas tata bahasa dan kelas bahasa**

Noam Chomsky mengklasifikasikan tata bahasa ke dalam beberapa kelas:

**1. Kelas tata bahasa regular** (*regular grammar*) atau tata bahasa tipe 3.

Aturan produksinya berbentuk:

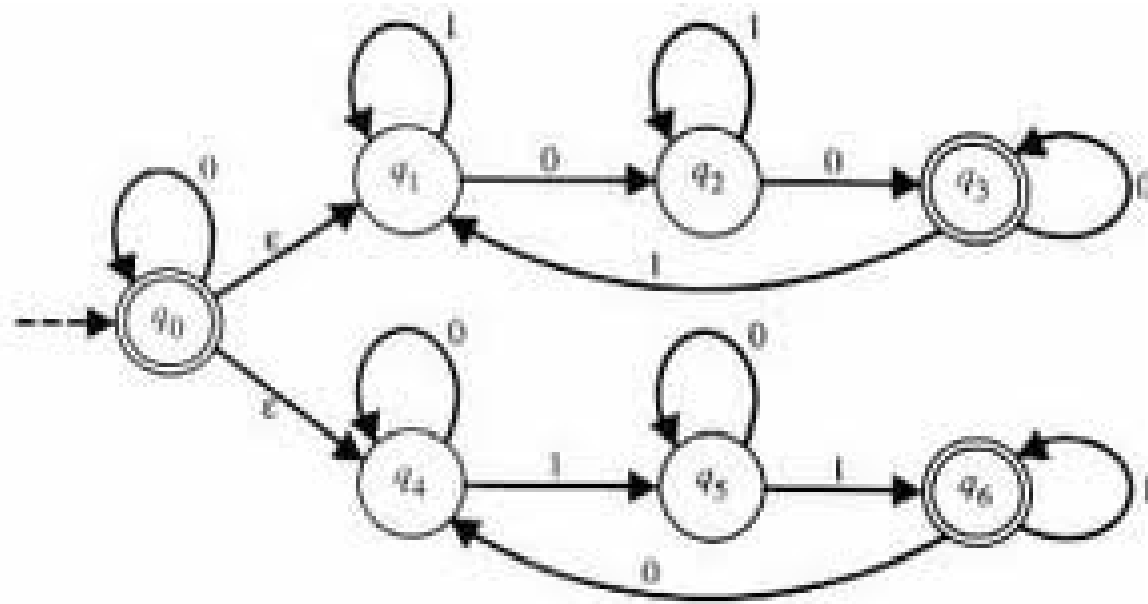
$$A \rightarrow a$$

$$A \rightarrow aB \quad (\text{atau } A \rightarrow Ba)$$

Bahasanya dinamakan **bahasa regular** (*regular language*)

Mesin yang mengenalinya adalah *Finite State Automaton* (FSA)





Contoh sebuah FSA

**2. Bahasa bebas konteks** (*context-free grammar*) atau tata bahasa tipe 2.

Aturan produksinya berbentuk:

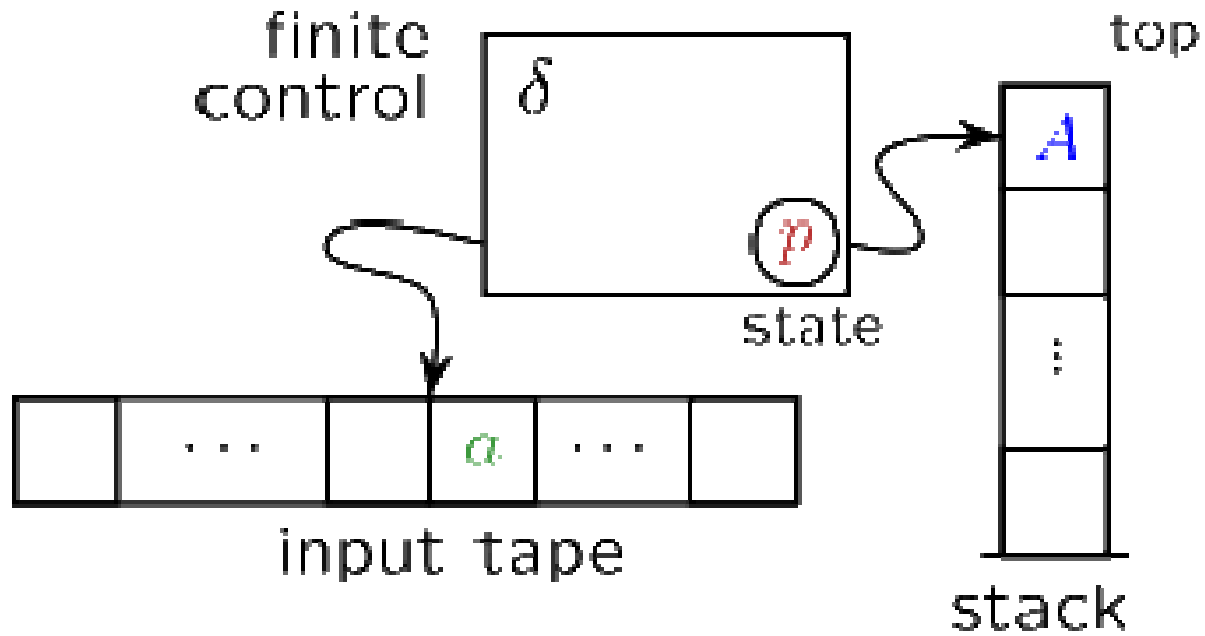
$$A \rightarrow \alpha$$

$\alpha$  adalah string yang dibentuk dari simbol terminal dan/atau simbol non terminal.

Bahasanya dinamakan **bahasa bebas-konteks** (*context-free language*) atau CFL.

Mesin yang mengenali bahasanya dinamakan *Push Down Automaton (PDA)*.

Tata bahasa tipe 3 termasuk di dalam tata bahasa tipe 2.



Contoh sebuah PDA

3. Kelas tata bahasa **peka-konteks** (*context-sensitive*) atau tata bahasa tipe 1.

Aturan produksinya berbentuk:

$$\alpha \rightarrow \beta$$

Panjang  $\beta$  selalu lebih besar atau sama dengan panjang  $\alpha$ .

Bahasanya dinamakan **bahasa peka-konteks** (*context-sensitive language*) atau CSL.

Mesin yang mengenali bahasanya dinamakan *Linear Bounded Automaton (LBA)*.

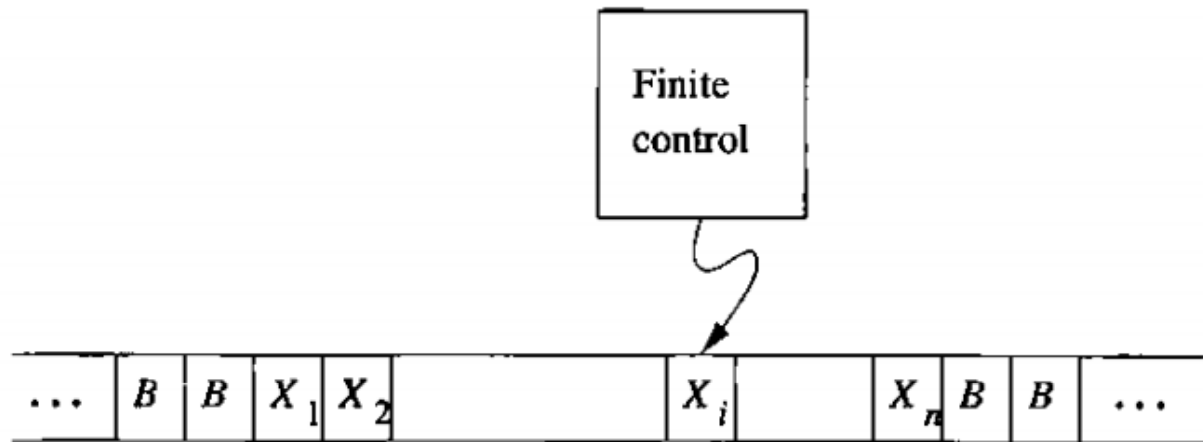
Tata bahasa tipe 3 dan tipe 2 termasuk di dalam tata bahasa tipe 1.

4. Tata bahasa **tanpa pembatasan** (*unrestricted grammar*) atau tata bahasa tipe 0.

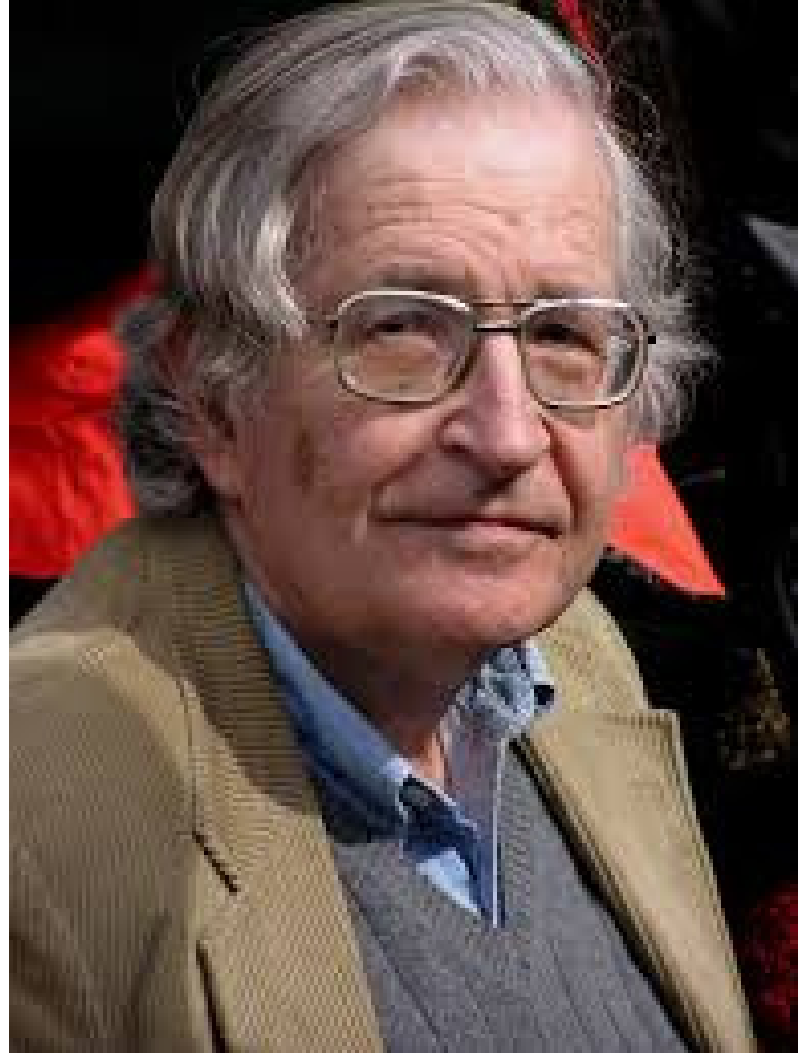
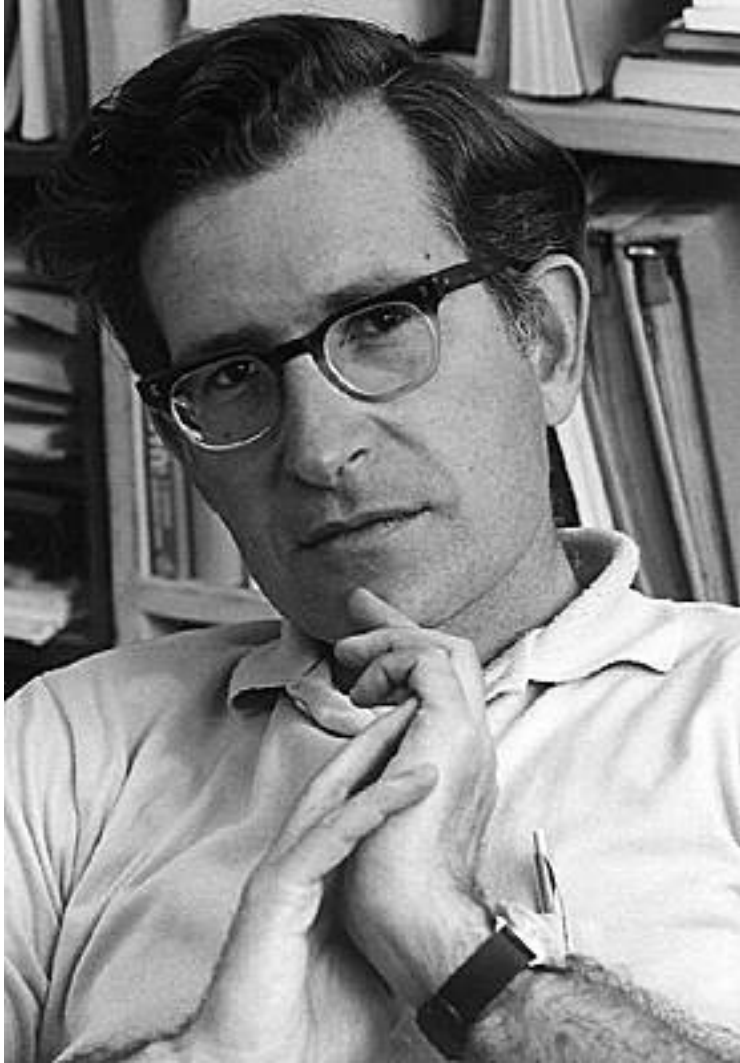
Aturan produksinya tidak mempunyai batasan, seperti pada tata bahasa tipe 3, 2, dan 1.

Bahasa yang dispesifikasikan oleh tata bahasa ini disebut **bahasa tanpa-pembatasan** (*unrestricted language*).

Mesin yang mengenali bahasa ini adalah **Mesin Turing** (*TuringMachine*)



Contoh Mesin Turing



Noam Chomsky

Tipe 0 (tanpa pembatasan)

Tipe 1  
(peka konteks)

Tipe 2  
(bebas konteks)

Tipe 3  
(regular)



<b>Tipe</b>	<b>Kelas Tata Bahasa</b>	<b>Mesin Pengenal Bahasa</b>
3	<i>Regular Grammar</i>	<i>Finite State Automaton (FSA)</i>
2	<i>Context-Free Grammar (CFG)</i>	<i>Push Down Automaton (PDA)</i>
1	<i>Context-Sensitive Grammar (CSG)</i>	<i>Linear Bounded Automaton</i>
0	<i>Unrestricted Grammar</i>	<i>Turing Machine</i>