

Operasi Dasar String (lanj), Grammar dan Hirarki Chomsky)

(Teori Bahasa dan Automata)

Operasi Dasar String

Diberikan string : $x = abc$

- Prefik string w adalah string yang dihasilkan dari string w dengan menghilangkan nol atau lebih simbol-simbol paling belakang dari string w tersebut.

Contoh : abc , ab , a , dan ε adalah semua $Prefix(x)$

- ProperPrefix string w adalah string yang dihasilkan dari string w dengan menghilangkan satu atau lebih simbol-simbol paling belakang dari string w tersebut.

Contoh : ab , a , dan ε adalah semua $ProperPrefix(x)$

String(x)	abc
Prefix(x)	a b c a b a \emptyset
ProperPrefix(x)	a b a \emptyset

Operasi Dasar String

Diberikan string : $x = abc$

- Postfix (atau Sufix) string w adalah string yang dihasilkan dari string w dengan menghilangkan nol atau lebih simbol-simbol paling depan dari string w tersebut.

Contoh : abc , bc , c , dan ε adalah semua $Postfix(x)$

- ProperPostfix (atau PoperSufix) string w adalah string yang dihasilkan dari string w dengan menghilangkan satu atau lebih simbol-simbol paling depan dari string w tersebut.

Contoh : bc , c , dan ε adalah semua $ProperPostfix(x)$

String(x)	abc
Postfix(x)	a b c b c c \emptyset
ProperPostfix(x)	b c c \emptyset

Operasi Dasar String

Diberikan string : $x = abc$

- Head string w adalah simbol paling depan dari string w .

Contoh : a adalah $Head(x)$

- Tail string w adalah string yang dihasilkan dari string w dengan menghilangkan simbol paling depan dari string w tersebut.

Contoh : bc adalah $Tail(x)$

Operasi Dasar String

- Substring string w adalah string yang dihasilkan dari string w dengan menghilangkan nol atau lebih simbol-simbol paling depan dan/atau simbol-simbol paling belakang dari string w tersebut.

Contoh : abc , ab , bc , a , b , c , dan ε adalah semua Substring(x)

- ProperSubstring string w adalah string yang dihasilkan dari string w dengan menghilangkan satu atau lebih simbol-simbol paling depan dan/atau simbol-simbol paling belakang dari string w tersebut.

Contoh : ab , bc , a , b , c , dan ε adalah semua Substring(x)

Operasi Dasar String

- Subsequence string w adalah string yang dihasilkan dari string w dengan menghilangkan nol atau lebih simbol-simbol dari string w tersebut.

Contoh : abc, ab, bc, ac, a, b, c , dan ε adalah semua $Subsequence(x)$

- ProperSubsequence string w adalah string yang dihasilkan dari string w dengan menghilangkan satu atau lebih simbol-simbol dari string w tersebut.

Contoh : ab, bc, ac, a, b, c , dan ε adalah semua $Subsequence(x)$

- Concatenation adalah penyambungan dua buah string. Operator concatenation adalah $concat$ atau tanpa lambang apapun.

Contoh : $concat(xy) = xy = abc123$

- Alternation adalah pilihan satu di antara dua buah string. Operator alternation adalah $alternate$ atau $|$

Contoh : $alternate(xy) = x|y = abc \text{ atau } 123$

Grammar

Tata bahasa (grammar) didefinisikan sebagai kumpulan dari himpunan-himpunan variabel, simbol-simbol terminal, simbol awal, yang dibatasi oleh aturan-aturan produksi.

Tahun 1959 Noam Chomsky melakukan penggolongan tingkatan bahasa menjadi 4 dan disebut Hirarki Chomsky.

Bahasa	Mesin Otomata	Batasan aturan produksi $\alpha \rightarrow \beta$
Regular / Tipe 3	Finite State Automata Meliputi: Deterministic Finite Automata dan Nondeterministic Finite Automata	α adalah sebuah simbol variabel. β Maksimal memiliki sebuah simbol variabel yang bila ada terletak pada posisi paling kanan
Bebas Konteks (Context Free)/Tipe 2	Push Down Automata (PDA)	α adalah sebuah simbol variabel
Context Sensitive /Tipe 1	Linear Bounded Automata	$ \alpha \leq \beta $
Unrestricted /Structure/ Natural Language/ Tipe 0	Mesin Turing	Tidak ada batasan

Tabel : Hirarki Chomsky

- Dalam pembicaraan grammar, anggota alfabet dinamakan simbol terminal atau token.
- Kalimat adalah deretan hingga simbol-simbol terminal.
- Bahasa adalah himpunan kalimat-kalimat dan anggota bahasa bisa tak hingga kalimat.
- Simbol terminal adalah simbol yang tidak dapat diturunkan lagi. Simbol non-terminal adalah simbol yang masih dapat diturunkan.
- Simbol terminal biasanya menggunakan huruf kecil, seperti, a, b,c,.....
- Simbol non-terminal biasanya menggunakan huruf besar seperti A,B, C, ...
- Simbol-simbol berikut adalah simbol terminal:
 - huruf kecil awal alfabet, misalnya: *a*, *b*, *c*
 - simbol operator, misalnya: +, −, dan ×
 - simbol tanda baca, misalnya: (,), dan ;
 - string yang tercetak tebal, misalnya: *if*, *then*, dan *else*.

- Simbol-simbol berikut adalah simbol non terminal:
 - huruf besar awal alfabet, misalnya: A, B, C ;
 - huruf S sebagai simbol awal;
 - string yang tercetak miring, misalnya: *expr* dan *stmt*.
- Huruf yunani melambangkan string yang tersusun atas simbol-simbol terminal atau simbol-simbol non terminal atau campuran keduanya, misalnya : α, β, γ

- Aturan produksi dinyatakan dalam bentuk $\alpha \rightarrow \beta$, (dibaca, α menghasilkan β). artinya : dalam sebuah derivasi dapat dilakukan penggantian simbol α dengan simbol β .
- α menyatakan simbol-simbol pada ruas kiri aturan produksi (sebelah kiri tanda \rightarrow).
- β menyatakan simbol-simbol pada ruas kanan aturan produksi (sebelah kanan tanda \rightarrow). Simbol-simbol pada aturan produksi dapat berupa simbol terminal atau simbol nonterminal/variabel.
- Derivasi adalah proses pembentukan sebuah kalimat atau sentensial. Sebuah derivasi dilambangkan sebagai : $\alpha \rightarrow \beta$.
- Sentensial adalah string yang tersusun atas simbol-simbol terminal atau simbol-simbol non terminal atau campuran keduanya.
- Kalimat adalah string yang tersusun atas simbol-simbol terminal. Kalimat adalah merupakan sentensial, sebaliknya belum tentu.

Grammar :

Grammar G didefinisikan sebagai pasangan 4 tuple : V_T , V_N , S , dan P , dan dituliskan sebagai $G(V_T , V_N , S, P)$

dimana :

V_T : himpunan simbol-simbol terminal (alfabet) \rightarrow kamus

V_N : himpunan simbol-simbol non terminal

$S \in V_N$: simbol awal (atau simbol start)

P : himpunan produksi

Contoh :

1. $G_1 : V_T = \{\mathbf{I, Love, Miss, You}\}, V_N = \{S, A, B, C\},$
 $P = \{S \rightarrow ABC, A \rightarrow \mathbf{I}, B \rightarrow \mathbf{Love} \mid \mathbf{Miss}, C \rightarrow \mathbf{You}\}$
 $S \Rightarrow ABC$
 $\Rightarrow \mathbf{IloveYou}$
 $L(G_1) = \{\mathbf{IloveYou, IMissYou}\}$

2. $G_2 : V_T = \{a\}, V_N = \{S\}, P = \{S \rightarrow aS \mid a\}$
 $S \Rightarrow aS$
 $\Rightarrow aaS$
 $\Rightarrow aaa$
 $L(G_2) = \{a^n \mid n \geq 1\}$
 $L(G_2) = \{\mathbf{a, aa, aaa, aaaa, \dots}\}$

Aturan produksi:

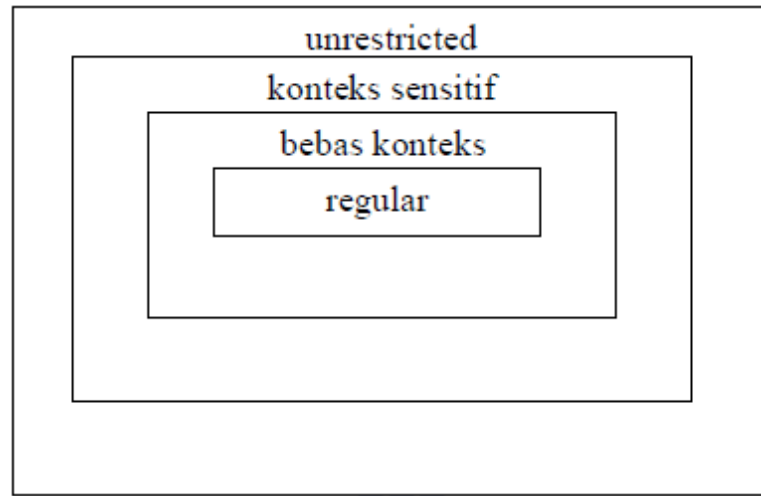
- $T \rightarrow a$ (dibaca “T menghasilkan a”)
- $E \rightarrow T | T+E$ (dibaca “E menghasilkan T atau E menghasilkan T + E”)
- Simbol “|” dibaca ‘atau’; digunakan untuk mempersingkat aturan produksi yang mempunyai ruas kiri yang sama.

Jadi penulisan aturan produksi :

- $E \rightarrow T | T+E$

adalah singkatan dari dua buah aturan produksi, yaitu:

- $E \rightarrow T$
- $E \rightarrow T+E$



Bahasa	Contoh Aturan Produksi
Regular / Tipe 3	$A \rightarrow def$ $A \rightarrow bc$ $A \rightarrow bcdE$ $C \rightarrow D$
Bebas Konteks (Context Free)/Tipe 2	$B \rightarrow CDeFg$ $D \rightarrow BcDe$
Context Sensitive /Tipe 1	$D \rightarrow ef \ (D < ef)$ $E \rightarrow \lambda \text{ (pengecualian)}$
Unrestricted /Structure/ Natural Language/ Tipe 0	$Abc \rightarrow deF$

Berdasarkan komposisi bentuk ruas kiri dan ruas kanan produksinya ($\alpha \rightarrow \beta$), Noam Chomsky mengklasifikasikan 4 tipe grammar :

1. Grammar tipe ke-0 : Unrestricted Grammar (UG)
Ciri : $\alpha, \beta \in (V_T | V_N)^*, |\alpha| > 0$
2. Grammar tipe ke-1 : Context Sensitive Grammar (CSG)
Ciri : $\alpha, \beta \in (V_T | V_N)^*, 0 < |\alpha| \leq |\beta|$
3. Grammar tipe ke-2 : Context Free Grammar (CFG)
Ciri : $\alpha \in V_N, \beta \in (V_T | V_N)^*$
4. Grammar tipe ke-3 : Regular Grammar (RG)
Ciri : $\alpha \in V_N, \beta \in \{V_T, V_T | V_N\}$ atau $\alpha \in V_N, \beta \in \{V_T, V_N | V_T\}$

Tipe sebuah grammar (atau bahasa) ditentukan dengan aturan sebagai berikut :

A language is said to be type- i ($i = 0, 1, 2, 3$) language if it can be specified by a type- i grammar but can't be specified any type- $(i+1)$ grammar.

Contoh Analisa Penentuan Type Grammar

1. Grammar G_1 dengan $P_1 = \{S \rightarrow aB, B \rightarrow bB, B \rightarrow b\}$.

Ruas kiri semua produksinya terdiri dari sebuah V_N maka G_1 kemungkinan tipe CFG atau RG.

Selanjutnya karena semua ruas kanannya terdiri dari sebuah V_T atau string V_TV_N maka G_1 adalah RG(3).

2. Grammar G_2 dengan $P_2 = \{S \rightarrow Ba, B \rightarrow Bb, B \rightarrow b\}$.

Ruas kiri semua produksinya terdiri dari sebuah V_N maka G_2 kemungkinan tipe CFG atau RG.

Selanjutnya karena semua ruas kanannya terdiri dari sebuah V_T atau string V_NV_T maka G_2 adalah RG(3).

Contoh Analisa Penentuan Type Grammar

3. Grammar G_3 dengan $P_3 = \{S \rightarrow Ba, B \rightarrow bB, B \rightarrow b\}$.

Ruas kiri semua produksinya terdiri dari sebuah V_N maka G_3 kemungkinan tipe CFG atau RG.

Selanjutnya karena ruas kanannya mengandung string V_TV_N (yaitu bB) dan juga string V_NV_T (Ba) maka G_3 bukan RG, dengan kata lain G_3 adalah CFG(2).

4. Grammar G_4 dengan $P_4 = \{S \rightarrow aAb, B \rightarrow aB\}$.

Ruas kiri semua produksinya terdiri dari sebuah V_N maka G_4 kemungkinan tipe CFG atau RG.

Selanjutnya karena ruas kanannya mengandung string yang panjangnya lebih dari 2 (yaitu aAb) maka G_4 bukan RG, dengan kata lain G_4 adalah CFG.

5. Grammar G_5 dengan $P_5 = \{S \rightarrow aA, S \rightarrow aB, aAb \rightarrow aBCb\}$.

Ruas kirinya mengandung string yang panjangnya lebih dari 1 (yaitu aAb) maka G_5 kemungkinan tipe CSG atau UG.

Selanjutnya karena semua ruas kirinya lebih pendek atau sama dengan ruas kananya maka G_5 adalah CSG.

6. Grammar G_6 dengan $P_6 = \{aS \rightarrow ab, SAc \rightarrow bc\}$.

Ruas kirinya mengandung string yang panjangnya lebih dari 1 maka G_6 kemungkinan tipe CSG atau UG.

Selanjutnya karena terdapat ruas kirinya yang lebih panjang daripada ruas kanannya (yaitu SAc) maka G_6 adalah UG.

Derivasi Kalimat dan Penentuan Bahasa

Tentukan bahasa dari masing-masing gramat berikut :

1. G_1 dengan $P_1 = \{1. S \rightarrow aAa, 2. A \rightarrow aAa, 3. A \rightarrow b\}$.

Jawab :

Derivasi kalimat terpendek :

$$S \Rightarrow aAa \quad (1)$$

$$\Rightarrow aba \quad (3)$$

Derivasi kalimat umum :

$$S \Rightarrow aAa \quad (1)$$

$$\Rightarrow aaAaa \quad (2)$$

...

$$\Rightarrow a^n A a^n \quad (2)$$

$$\Rightarrow a^n b a^n \quad (3)$$

Dari pola kedua kalimat disimpulkan : $L(G) = \{ aba \mid n \geq 1 \}$

2. G_2 dengan :

$P = \{1. S \rightarrow aS, 2. S \rightarrow aB, 3. B \rightarrow bC, 4. C \rightarrow aC, 5. C \rightarrow a\}$.

Jawab :

Derivasi kalimat terpendek :

$$S \Rightarrow aB \quad (2)$$

$$\Rightarrow abC \quad (3)$$

$$\Rightarrow aba \quad (5)$$

Derivasi kalimat umum :

$$S \Rightarrow aS \quad (1)$$

...

$$\Rightarrow a^{n-1}S \quad (1)$$

$$\Rightarrow a^n B \quad (2)$$

$$\Rightarrow a^n b C \quad (3)$$

$$\Rightarrow a^n b a C \quad (4)$$

...

$$\Rightarrow a^n b a^{m-1} C \quad (4)$$

$$\Rightarrow a^n b a^m \quad (5)$$

Dari pola kedua kalimat disimpulkan : $L_2(G_2) = \{ a^n b a^m \mid n \geq 1, m \geq 1 \}$

Tentukan bahasa dari masing-masing gramat berikut :

3. G_3 dengan

$P_3 = \{1. S \rightarrow aSBC, 2. S \rightarrow abC, 3. bB \rightarrow bb, 4. bC \rightarrow bc, 5. CB \rightarrow BC, 6. cC \rightarrow cc\}$.

Jawab :

Derivasi kalimat terpendek 1:

$S \Rightarrow abC$ (2)

$\Rightarrow abc$ (4)

Derivasi kalimat terpendek 2 :

$S \Rightarrow aSBC$ (1)

$\Rightarrow aabCBC$ (2)

$\Rightarrow aabBCC$ (5) $aabcBC$ (4)

$\Rightarrow aabbCC$ (3)

$\Rightarrow aabbcC$ (4)

$\Rightarrow aabbcc$ (6)

Derivasi kalimat terpendek 3 :

$S \Rightarrow aSBC$ (1)

$\Rightarrow aaSBCBC$ (1)

$\Rightarrow aaabCBCBC$ (2)

$\Rightarrow aaabBCCBC$ (5)

$\Rightarrow aaabBCBCC$ (5)

$\Rightarrow aaabBBCCC$ (5)

$\Rightarrow aaabbBCCC$ (3)

$\Rightarrow aaabbbCCC$ (3)

$\Rightarrow aaabbbcCC$ (4)

$\Rightarrow aaabbbccC$ (6)

$\Rightarrow aaabbbccc$ (6)

► Dari pola ketiga kalimat disimpulkan : $L_3(G_3) = \{ a^n b^n c^n \mid n \geq 1 \}$

Menentukan Grammar Sebuah Bahasa

1. Tentukan sebuah grammar regular untuk bahasa $L = \{ a \mid n \geq 1 \}$

Jawab :

$$P_1(L_1) = \{ S \rightarrow aS \mid a \}$$

2. Tentukan sebuah grammar bebas konteks untuk bahasa :

L_2 : himpunan bilangan bulat non negatif ganjil

Jawab :

Langkah kunci : digit terakhir bilangan harus ganjil.

$$V_t = \{0, 1, 2, \dots, 9\}$$

$$V_n = \{S, G, J\}$$

$$P = \{ S \rightarrow HT \mid JT \mid J; \quad T \rightarrow GT \mid JT \mid J; \quad H \rightarrow 2 \mid 4 \mid 6 \mid 8; \quad G \rightarrow 0 \mid 2 \mid 4 \mid 6 \mid 8; \quad J \rightarrow 1 \mid 3 \mid 5 \mid 7 \mid 9 \}$$

$$P = \{ S \rightarrow GS \mid JS \mid J; \quad G \rightarrow 0 \mid 2 \mid 4 \mid 6 \mid 8; \quad J \rightarrow 1 \mid 3 \mid 5 \mid 7 \mid 9 \}$$

Buat dua buah himpunan bilangan terpisah : genap (G) dan ganjil (J)

$$P_2(L_2) = \{ S \rightarrow J \mid GS \mid JS, \quad G \rightarrow 0 \mid 2 \mid 4 \mid 6 \mid 8, \quad J \rightarrow 1 \mid 3 \mid 5 \mid 7 \mid 9 \}$$

3. Tentukan sebuah grammar bebas konteks untuk bahasa :

$B.L_3$ = himpunan semua identifier yang sah menurut bahasa pemrograman Pascal dengan batasan : terdiri dari simbol huruf kecil dan angka, panjang identifier boleh lebih dari 8 karakter

Jawab :

Langkah kunci : karakter pertama identifier harus huruf.

Buat dua himpunan bilangan terpisah : huruf (H) dan angka (A)

$S \rightarrow HT|H;$ $T \rightarrow HT|AT|H|A;$ $H \rightarrow a|..|z;$ $A \rightarrow 0|..|9$

$P_3(L_3) = \{S \rightarrow H | HT, T \rightarrow AT | HT | H | A,$
 $H \rightarrow a | b | c | \dots, A \rightarrow 0 | 1 | 2 | \dots\}$

4. Tentukan sebuah grammar bebas konteks untuk bahasa :

L_4 = bilangan bulat non negatif genap. Jika bilangan tersebut terdiri dari dua digit atau lebih maka nol tidak boleh muncul sebagai digit pertama.

Jawab :

Langkah kunci : Digit terakhir bilangan harus genap. Digit pertama tidak boleh nol. Buat tiga himpunan terpisah : bilangan genap tanpa nol (G), bilangan genap dengan nol (N), serta bilangan ganjil (J).

$P_4(L_4) = \{S \rightarrow N \mid GA \mid JA, A \rightarrow N \mid NA \mid JA, G \rightarrow 2 \mid 4 \mid 6 \mid 8,$
 $N \rightarrow 0 \mid 2 \mid 4 \mid 6 \mid 8, J \rightarrow 1 \mid 3 \mid 5 \mid 7 \mid 9\}$