Teori Bahasa & Otomata

Muhamad Nursalman

Pendikom/Ilkom

Universitas Pendidikan Indonesia

Daftar Isi

- Bab 1 Pendahuluan
- Bab 2 Matematika Dasar
- Bab 3 Dasar-Dasar Teori Bahasa
- Bab 4 Representasi Bahasa
- Bab 5 Klasifikasi Grammar Noam Chomsky

Bab 1 Pendahuluan

- **Komponen Ilmu Informatika**
 - Ide & model fundamental yang mendasari komputasi
 - Teknik rekayasa untuk perancangan sistem komputasi

Bab1 Pendahuluan (Model Komputasi)

- Finite automata/finite state automata (FSA)
 - Deterministic finite automata (DFA)
 - Non deterministic finite automata (NDFFA)
- Pushdown automata (PA)
 - Deterministic pushdown automata (DPA)
 - Non deterministic pushdown automata (NDPA)
- Turing machine (TM)

Bab 1 Pendahuluan (Teori Komputasi)

- Apa yang dimaksud dengan mengkomputasi?
- Apa yang dapat dikomputasi?
- Seberapa kompleks untuk mengkomputasi sesuatu?

Bab 2 Matematika Dasar (Himpunan)

- Himpunan bagian, $A \subseteq B$
- Penggabungan, $A \cup B$
- Irisan, $A \cap B$
- Complement (Relative/Absolute)
- Cartesian Product,

$$A \times B = \{(x, y) \mid (x \in A) \text{ dan } (y \in B)\}$$

Bab2 Matematika Dasar (Relasi)

Sifat-sifat relasi:

- Reflexive, $\forall x \in X, xRx \Rightarrow (x, x) \in R$
- Symmetric, $x, y \in X, xRy \Rightarrow yRx$
- Transitive, $x, y, z \in X, xRy \ \& \ yRz \Rightarrow xRz$
- Irreflexive, $\forall x \in X \Rightarrow (x, x) \notin R$
- Antisymmetric $x, y \in X, xRy \ \& \ yRx \Rightarrow x = y$

Bab 2 Matematika Dasar (Relasi)

Transitive Closure

- Definisi: Bila X adalah suatu himpunan berhingga dan R adalah relasi pada X .
Relasi $R^+ = R \cup R^2 \cup R^3 \dots$ pada X , disebut transitive closure R pada X .

Bab 2 Matematika Dasar (Relasi)

- Transitive closure R^+ relasi R pada suatu himpunan berhingga X adalah transitif. Juga untuk suatu relasi transitif P lain pada X dimana $R \subseteq P$, kita mempunyai $R^+ \subseteq P$. Dalam arti ini, R^+ adalah relasi transitif terkecil yang berisi R .

Bab 2 Matematika Dasar (Logika)

$$\neg(\neg P) = P$$

$$(P \vee Q) = (\neg P \Rightarrow Q)$$

Hukum de Morgan

$$\neg(P \vee Q) = (\neg P \wedge \neg Q)$$

$$\neg(P \wedge Q) = (\neg P \vee \neg Q)$$

Hukum Distributif

$$P \vee (Q \wedge R) = (P \vee Q) \wedge (P \vee R)$$

$$P \wedge (Q \vee R) = (P \wedge Q) \vee (P \wedge R)$$

Bab 2 Matematika Dasar (Logika)

Hukum Komutatif

$$(P \wedge Q) = (Q \wedge P)$$

$$(P \vee Q) = (Q \vee P)$$

Hukum Asosiatif

$$((P \wedge Q) \wedge R) = (P \wedge (Q \wedge R))$$

$$((P \vee Q) \vee R) = (P \vee (Q \vee R))$$

Hukum Kontrapositif

$$(P \Rightarrow Q) = (\neg Q \Rightarrow \neg P)$$

Bab 2 Matematika Dasar (Graph)

- Dua graph disebut ekivalen (isomorphic) jika keduanya berperilaku identik menurut kriteria-kriteria graph.
- Syarat perlu dua graph adalah isomorphic:
 - Jumlah simpul ke-2 graph sama
 - Jumlah busur ke-2 graph sama
 - Jumlah simpul yang sama dengan derajat yang diberikan

Bab 2 Matematika Dasar (Graph)

Pohon adalah graph G dengan n simpul, jika:

1. G terhubung dan tanpa sirkit, atau
2. G terhubung dan $n-1$ busur, atau
3. G tanpa sirkit dan mempunyai $n-1$ busur, atau
4. Terdapat tepat satu path di antara pasangan simpul di G , atau
5. G adalah graph terhubung minimal

Bab 2 Matematika Dasar (Graph)

- Terdapat beragam algoritma penentuan graph terhubung
 1. Algoritma permutasi baris dan kolom matriks
 2. Algoritma memanfaatkan DFT dan BFT
 3. Algoritma menggunakan operasi fusion
 4. Algoritma warshall

Bab 3 Dasar-Dasar Teori Bahasa

- Penyambungan [o]
 - ‘a’ o ‘b’ = ‘ab’
- String pada alphabet V
 - $V = \{ 'a', 'b', 'c', 'd' \}; 'a', 'abcd', 'bbba'$
 - $V^n = V o V o \dots o V$
 - $V^+ = V^1 \cup V^2 \cup V^3 \cup \dots$
 - $V^* = \{ \varepsilon \} \cup V^+$, ε adalah string kosong dan mempunyai sifat identitas.

Bab 3 Dasar-Dasar Teori Bahasa

- Definisi: Diberikan alphabet V , bila $x = a_1a_2\dots a_n$ dan $y = b_1b_2\dots b_m$ adalah string pada V , maka x dan y adalah sama jika dan hanya jika $n=m$ dan untuk masing-masing $i = 1, 2, \dots, n$, $a_i = b_i$.
- Bahasa:
 - Subset L dari V^* disebut bahasa pada V .Contoh: $V^*, \emptyset, \{\varepsilon\}$

Bab 3 Dasar-Dasar Teori Bahasa

- Terapan (Bahasa Pascal)
 - Aspek Leksik
 - Alfabeta Pascal digunakan untuk membentuk token yang berupa keyword dan identifier.
 - Aspek Sintaks
 - Penyambungan token-token yang memenuhi syarat sintaks Pascal.
 - Aspek Semantiks
 - Setelah memenuhi aspek leksik dan sintaks, maka untuk menjadi program Pascal juga harus memenuhi aspek semantiksnya.

Bab 3 Dasar-Dasar Teori Bahasa

- Definisi: String pada alphabet V_T adalah
 1. ε adalah string pada V_T
 2. Jika x adalah string pada V_T dan a adalah elemen V_T , maka xa adalah string V_T .
 3. y adalah string pada V_T jika dan hanya jika mengikuti aturan (1) dan (2).
- Jika x dan y adalah string, maka string xy adalah penyambungan x dan y .

Bab 3 Dasar-Dasar Teori Bahasa

- Definisi: V_T^* menunjukkan himpunan berisi semua string pada V_T termasuk ε . Dengan demikian bahasa L adalah $L \subset V_T^*$.
- Himpunan kosong $\emptyset$, adalah bahasa.
Himpunan $\{\varepsilon\}$ adalah bahasa yang hanya berisi string kosong.
- $\emptyset$ dan $\{\varepsilon\}$ adalah dua bahasa yang berbeda.

Bab 3 Dasar-Dasar Teori Bahasa

- Definisi: Jika L_1 bahasa pada alphabet V_{T_1} dan L_2 bahasa pada alphabet V_{T_2} . Maka L_1L_2 disebut penyambungan (*concatenation*) atau perkalian (*product*) dari L_1 dan L_2 yaitu bahasa dengan $\{xy \mid x \in L_1 \text{ dan } y \in L_2\}$

Bab 3 Dasar-Dasar Teori Bahasa

- Definisi: Ketertutupan (*Closure*) L , ditandai dengan L^* didefinisikan sebagai berikut:
 1. $L^0 = \{e\}$
 2. $L^n = LL^{n-1}$ untuk $n \geq 1$
 3. $L^* = \bigcup_{n \geq 0} L^n$
 4. $L^+ = \bigcup_{n \geq 1} L^n$
 5. $L = L^+ \cup \{e\}$

Bab 3 Dasar-Dasar Teori Bahasa

- Union L dan M ditulis dengan $L \cup M$
 - Adalah $\{s \mid s \in L \text{ atau } s \in M\}$
- Penyambungan L dan M ditulis dengan LM
 - Adalah $\{st \mid s \in L \text{ dan } t \in M\}$
- Kleene Closure dari L ditulis L^*
 - Adalah $\cup_{i=0}^{\infty} L$
- Positive Closure dari L ditulis L^+
 - Adalah $\cup_{i=1}^{\infty} L$

Bab 3 Dasar-Dasar Teori Bahasa

Homomorphism

- Definisi: Bila V_{T_1} dan V_{T_2} alphabet, maka homomorphism adalah pemetaan $h : V_{T_1} \rightarrow V_{T_2}^*$
Kita memperluas domain homomorphism h ke $V_{T_1}^*$ dengan $h(e) = e$ dan $h(x)h(a)$ untuk semua x dalam $V_{T_1}^*$, a dalam V_{T_1} .

Bab 3 Dasar-Dasar Teori Bahasa

- Definisi: Jika $h : V_{T_1} \rightarrow V_{T_2}^*$ adalah homomorphism, maka relasi $h^{-1} : V_{T_2}^* \rightarrow p(V_{T_1})$ yang didefinisikan di bawah ini disebut inverse homomorphism.
- Secara formal:
$$h^{-1}(L) = \{y \in L \mid \exists x \text{ s.t. } h(x) = y\}$$

Bab 4 Representasi Bahasa

- Definisi: Grammar adalah sistem matematis untuk mendefinisikan bahasa. Bahasa yang didefinisikan oleh grammar adalah himpunan string yang hanya berisi terminal dan dapat diturunkan mulai dari simbol tertentu yang dikhususkan yang disebut S atau simbol mula (*starting symbol*).

Bab 4 Representasi Bahasa

- *Grammar* didefinisikan oleh 4 tupel

$G = (V_N, V_T, S, \phi)$, dimana V_N adalah himpunan simbol non-terminal, S adalah sebuah elemen dari V_N yang khusus yang disebut dengan simbol awal. Dan ϕ adalah himpunan bagian tak kosong dari relasi dari $(V_T \cup V_N)^* V_N (V_T \cup V_N)^*$ ke $(V_T \cup V_N)^*$. Secara umum dapat ditulis (α, β) yang disebut aturan produksi atau aturan penulisan kembali.

Bab 4 Representasi Bahasa

- Definisi (Penurunan Langsung): Bila $G = (V_N, V_T, S, \phi)$ adalah grammar. Untuk $\sigma, \psi \in V^*$, σ dikatakan penurunan langsung dari ψ ditulis dengan $\psi \Rightarrow \sigma$, jika terdapat string ϕ_1 dan ϕ_2 (termasuk string kosong) sehingga $\psi = \phi_1 \alpha \phi_2$ dan $\sigma = \phi_1 \beta \phi_2$ dan $\alpha \rightarrow \beta$ merupakan produksi dari G .

Bab 4 Representasi Bahasa

- Bentuk Kalimat

Bentuk kalimat (sentential form) adalah tiap penurunan nonterminal S unik. Bahasa L yang dihasilkan grammar G adalah kumpulan semua bentuk kalimat yang simbol-simbolnya adalah simbol terminal.

$$L(G) = \{ \sigma \mid S \xRightarrow{*} \sigma \text{ dan } \sigma \in V_T \}$$

Bab 4 Representasi Bahasa

- Bahasa yang didefinisikan oleh recognizer adalah himpunan string masukan yang diterimanya. Karakteristik bahasa yang diterima recognizer adalah:
 1. Bahasa L adalah right linear jika dan hanya jika L didefinisikan oleh finite automaton searah deterministik.
 2. Bahasa L adalah context free jika dan hanya jika L didefinisikan pushdown automaton searah nondeterministik.

Bab 4 Representasi Bahasa

3. Bahasa L adalah context sensitive jika dan hanya jika L didefinisikan oleh pushdown bounded automaton linear dua arah non deterministik.
4. Bahasa yang secara rekursif terdaftar jika dan hanya jika L didefinisikan oleh mesin turing.

Bab 4 Representasi Bahasa

- **Translasi Bahasa** Translasi adalah himpunan string. Kompilator mendefinisikan translasi sebagai pasangan. Jika kita anggap kompilator berisi 3 tahap, yaitu analisis leksik, sintaks, dan pembangkitan kode, maka masing-masing tahap itu mendefinisikan translasi.

Bab 4 Representasi Bahasa

- Analisis Leksik adalah translasi string-string yang merepresentasikan program sumber dipetakan menjadi string-string token.
- Analisis Sintaks memetakan string-string token menjadi string-string yang merepresentasikan pohon sintaks.
- Pembangkit kode kemudian mengambil string-string yang dihasilkan analisis sintaks menjadi bahasa mesin atau assembly.

Bab 4 Representasi Bahasa

- Definisi (Translasi): Misalkan V_T adalah alphabet masukan dan Δ alphabet keluaran. Kita mendefinisikan translasi satu bahasa $L_1 \subseteq V_T^*$ ke bahasa $L_2 \subseteq \Delta^*$ sebagai relasi T dari V_T^* ke Δ^* di mana domain T adalah L_1 dan range T adalah L_2 .

Bab 4 Representasi Bahasa

- Contoh penulisan grammar lengkap:

$G = (V_N, V_T, S, \phi)$ dengan

$V_N = \{I, L, D\}$

$V_T = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

$S = I$

$\phi = \{I \rightarrow L, I \rightarrow ID, I \rightarrow IL, L \rightarrow a, L \rightarrow b, \dots, L \rightarrow z, D \rightarrow 0, D \rightarrow 1, \dots, D \rightarrow 9\}$

Bab 4 Representasi Bahasa

- Penulisan dengan BNF:

$\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle | \langle \text{identifier} \rangle \langle \text{letter} \rangle |$
 $\langle \text{identifier} \rangle \langle \text{digit} \rangle$

$\langle \text{letter} \rangle ::= a | b | c | \dots | z$

$\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$

Bab 5 Klasifikasi Grammar

Noam Chomsky

- Definisi: G dinyatakan sebagai
 1. Right linear jika tiap produksi pada P berbentuk $A \rightarrow xB$ atau $A \rightarrow A$, di mana A dan B adalah V_N dan x adalah V_T^* .
 2. Context free jika tiap produksi pada P berbentuk $A \rightarrow \alpha$, di mana A adalah V_N dan α adalah $(V_N \cup V_T)^*$
 3. Context sensitive jika tiap produksi P berbentuk $\alpha \rightarrow \beta$ di mana $|\alpha| \leq |\beta|$
 4. Grammar tanpa pembatasan-pembatasan di atas disebut unrestricted grammar.

Bab 5 Klasifikasi Grammar

Noam Chomsky

1. **Kelas 0** Unrestricted grammar (aturan produksinya tak dibatasi)
2. **Kelas 1** Context sensitive grammar, di mana $\alpha \rightarrow \beta$ dengan $|\alpha| \leq |\beta|$
3. **Kelas 2** Context free grammar, di mana $\alpha \rightarrow \beta$
 $\alpha \in V_N$ dan β adalah $(V_N \cup V_T)^*$
4. **Kelas 3** Regular grammar di mana $\alpha \rightarrow \beta$
dengan $|\alpha| \leq |\beta|$, $\alpha \in V_N$ dan β berbentuk aB atau a , dengan $a \in V_T^*$ dan $B \in V_N$.

Bab 5 Klasifikasi Grammar Noam Chomsky (Kelas 3: Regular Gramar)

Contoh:

$G = (\{S, A, B, C\}, \{a, b\}, \phi, S)$, ϕ adalah

$$S \rightarrow aS \mid aB$$

$$B \rightarrow bC$$

$$C \rightarrow aC \mid a$$

Bahasa yang dihasilkan adalah

$$L(G) = \{ a^m b a^n \mid m, n \geq 1 \}$$

Bab 5 Klasifikasi Grammar Noam Chomsky (Kelas 2: CFG)

Contoh:

$G = (\{S, A, B, C\}, \{a, b\}, \phi, S)$, ϕ adalah

$$S \rightarrow aSbb \mid abb$$

Bahasa yang dihasilkan adalah

$$L(G) = \{ a^n b^{2n} \mid n \geq 1 \}$$

Bab 5 Klasifikasi Grammar Noam Chomsky (Kelas 1: CSG)

Contoh:

$G = (\{S, A, B, C\}, \{a, b\}, \phi, S)$, ϕ adalah

$$S \rightarrow 0A1$$

$$0A \rightarrow 00A1$$

$$A \rightarrow 1$$

Bahasa yang dihasilkan adalah

$$L(G) = \{0^n 1^n \mid n \geq 1\}$$

Bab 5 Klasifikasi Grammar Noam Chomsky (Kelas 0: UG)

Contoh:

$G = (\{S, A, B, C\}, \{a, b\}, \phi, S)$, ϕ adalah

$S \rightarrow CD$ $Ab \rightarrow bA$

$C \rightarrow aCA \mid bCB$ $Ba \rightarrow aB$

$AD \rightarrow aD$ $Bb \rightarrow bB$

$BD \rightarrow bD$ $C \rightarrow e$

$Aa \rightarrow aA$ $D \rightarrow e$

Bahasa yang dihasilkan adalah $L(G) = \{ww \mid w \in \{a,b\}^*\}$. Bahasa yang dihasilkan grammar ini merupakan himpunan yang dikenali dengan mesin turing.