

MODUL X

TEORI BAHASA DAN AUTOMATA

Tujuan :

Mahasiswa memahami tentang tata bahasa bebas konteks dan membangun pohon penurunan tata bahasa bebas konteks

Materi :

Pohon Derivatif

- Tata Bahasa Bebas Konteks
- Parsing
- Ambiguitas

Penyederhanaan Tata Bebas Konteks

- Reduksi produksi useless
- Reduksi produksi unit
- Reduksi produksi ϵ

POHON PENURUNAN

Tata Bahasa Bebas Konteks

Bila pada tata bahasa regular terdapat pembatasan pada ruas kanan atau hasil produksinya, maka pada tata bahasa bebas konteks/ *context free grammar*, selanjutnya kita sebut CFG, tidak terdapat pembatasan hasil produksinya. Pada aturan produksi:

$$\alpha \rightarrow \beta$$

batasannya hanyalah ruas kiri (α) adalah sebuah symbol variabel.

Contoh aturan produksi yang termasuk CFG:

$$B \rightarrow CDeFg$$

$$D \rightarrow BcDe$$

Seperti halnya pada tata bahasa regular, sebuah tata bahasa bebas konteks adalah suatu cara yang menunjukkan bagaimana menghasilkan untai-untai dalam sebuah bahasa. Seperti kita ketahui, pada saat menurunkan suatu *string*, symbol-simbol variabel akan mewakili bagian-bagian yang belum yang belum terturunkan dari *string* tersebut. Bila pada tata bahasa regular, bagian yang belum terturunkan tersebut selalu terjadi pada suatu ujung, pada tata bahasa bebas konteks bisa terdapat lebih banyak bagian yang belum terturunkan itu dan bisa terjadi dimana saja. Ketika penurunan itu sudah lengkap, semua bagian yang belum terturunkan telah diganti oleh *string-string* (yang mungkin saja kosong) dari himpunan symbol terminal. Bahasa bebas konteks menjadi dasar dalam pembentukan suatu *parser*/proses analisis sintaksis. Bagian sintaks dalam suatu kompilator kebanyakan didefinisikan dalam tata bahasa bebas konteks.

Parsing

Sebuah pohon (*tree*) adalah suatu *graph* terhubung tidak sirkuler, yang memiliki satu simpul (*node*)/*vertex* disebut akar (*root*) dan dari situ memiliki lintasan ke setiap simpul. Gambar 1 memberikan contoh sebuah *tree* yang menguraikan kalimat dalam bahasa Inggris:

The quick brown fox jumped over the lazy dog

Pohon penurunan (*derivation tree*/parse tree) berguna untuk menggambarkan bagaimana memperoleh suatu *string* (untai) dengan cara menurunkan simbol-simbol variabel menjadi simbol-simbol terminal. Setiap simbol variabel akan diturunkan menjadi terminal, sampai tidak ada yang belum tergantikan.

Misal terdapat tata bahasa bebas konteks dengan aturan produksi (symbol awal S, selanjutnya didalam bab ini digunakan sebagai symbol awal untuk tata bahasa bebas konteks adalah S):

$$S \rightarrow AB$$

$$A \rightarrow aA \mid a$$

$$B \rightarrow bB \mid b$$

Akan kita gambarkan pohon penurunan untuk memperoleh untai: 'aabbb'. Pada pohon tersebut symbol awal akan menjadi akar (*root*). Setiap kali penurunan dipilih aturan produksi yang menuju ke solusi. Simbol-simbol variabel akan menjadi simpul-simpul yang mempunyai anak. Simpul-simpul yang tidak mempunyai anak akan menjadi symbol terminal. Kalau kita baca symbol terminal yang ada pada gambar 2 dari kiri ke kanan akan diperoleh unatai 'aabbb'.

Proses penurunan atau *parsing* bisa dilakukan dengan cara:

- Penurunan terkiri (*leftmost derivation*): symbol variabel terkiri yang diperluas terlebih dahulu.
- Penurunan terkanan (*rightmost derivation*): symbol variabel terkiri yang diperluas terlebih dahulu

Misal terdapat tata bahasa bebas konteks:

$$S \rightarrow aAS \mid a$$

$$A \rightarrow SbA \mid ba$$

Untuk memperoleh untai 'aabbaa' dari tata bahasa bebas konteks diatas (' \Rightarrow ' bisa dibaca 'menurunkan')

- Dengan penurunan terkiri: $S \Rightarrow aAS \Rightarrow aSbAS \Rightarrow aabAS \Rightarrow aabbaS \Rightarrow aabbaa$
- Dengan penurunan terkanan: $S \Rightarrow aAS \Rightarrow aAa \Rightarrow aSbAa \Rightarrow aSbbaa \Rightarrow aabbaa$

Kita lihat pohon penurunannya pada gambar 3. Meskipun proses penurunannya berbeda akan tetap memiliki pohon penurunan yang sama.

Biasanya persoalan yang diberikan berkaitan dengan pohon penurunan, adalah untuk mencari penurunan yang hasilnya menuju kepada suatu untai yang ditentukan. Dalam hal ini, perlu untuk melakukan percobaan pemilihan aturan produksi yang bisa menuju ke solusi. Misalkan sebuah tata bahasa bebas konteks memiliki aturan produksi:

$$S \rightarrow aB \mid bA$$

$$A \rightarrow a \mid aS \mid bAA$$

$$B \rightarrow b \mid bS \mid aBB$$

Pohon penurunan untuk memperoleh : 'aaabbabbaa' bisa dilihat pada gambar 4

AMBIGUITAS

Ambiguitas/kedwitarian terjadi bila terdapat lebih dari satu pohon penurunan yang berbeda untuk memperoleh suatu untai.

Misalkan terdapat tata bahasa bebas konteks:

$$S \rightarrow A \mid B$$

$$A \rightarrow a$$

$$B \rightarrow a$$

Untuk memperoleh untai 'a' bisa terdapat dua cara penurunan:

- $S \Rightarrow A \Rightarrow a$
- $S \Rightarrow B \Rightarrow a$

Contoh lain, terdapat tata bahasa bebas konteks:

$$S \rightarrow SbS \mid ScS \mid a$$

Kita bisa menurunkan untai 'abaca' dalam dua cara

- $S \Rightarrow SbS \Rightarrow SbScS \Rightarrow SbSca \Rightarrow Sbaca \Rightarrow abaca$
- $S \Rightarrow ScS \Rightarrow SbScS \Rightarrow abScS \Rightarrow abacS \Rightarrow abaca$

Pohon penurunannya bisa dilihat pada gambar 5 dan 6

Kita lihat untuk untai yang sama ('abaca') dapat dibuat pohon penurunan yang berbeda, maka dapat dikatakan tata bahasa bebas konteks tersebut ambigu. Jadi untuk menunjukkan bahwa suatu tata bahasa bebas konteks ambigu, bisa dilakukan dengan menemukan untai yang memungkinkan pembentukan lebih dari satu pohon penurunan. Ambiguitas dapat menimbulkan masalah pada bahasa-bahasa tertentu, baik pada bahasa alami maupun pada bahasa pemrograman. Bila suatu struktur bahasa memiliki lebih dari satu dekomposisi (penurunan), dan susunannya akan menentukan arti, maka artinya menjadi ambigu.

PENYEDERHANAAN TATA BAHASA BEBAS KONTEKS

Tujuan Penyederhanaan

Penyederhanaan tata bahasa bebas konteks bertujuan untuk melakukan pembatasan sehingga tidak menghasilkan pohon penurunan yang memiliki kerumitan

yang tak perlu atau aturan produksi yang tak berarti. Misalkan terdapat tata bahasa bebas konteks (dengan symbol awal S, dalam bab ini kita gunakan sebagai symbol awal untuk tata bahasa bebas konteks adalah S) :

$$S \rightarrow AB \mid a$$

$$A \rightarrow a$$

Kelemahan tata bahasa bebas konteks di atas, aturan produksi $S \rightarrow AB$ tidak berarti karena B tidak memiliki penurunan.

untuk tata bahasa bebas konteks berikut :

$$S \rightarrow A$$

$$A \rightarrow B$$

$$B \rightarrow C$$

$$C \rightarrow D$$

$$D \rightarrow a \mid A$$

memiliki kelemahan terlalu panjang jalannya padahal berujung pada $S \rightarrow a$, produksi $D \rightarrow A$ juga menyebabkan kerumitan.

Suatu tata bahasa bebas konteks dapat disederhanakan dengan melakukan :

1. penghilangan produksi useless
2. penghilangan produksi unit
3. penghilangan produksi ϵ

Selanjutnya akan kita bahas satu persatu cara penyederhanaan tersebut.

Penghilangan Produksi Useless

Disini produksi *useless* didefinisikan sebagai :

- Produksi yang memuat symbol variabel yang tidak memiliki penurunan yang akan menghasilkan terminal-terminal seluruhnya (kita sebut saja sebagai 'menuju terminal'), produksi ini tidak berguna karena bila diturunkan tidak akan pernah selesai (masih ada symbol variabel yang tersisa)
- produksi yang tidak akan pernah dicapai dengan penurunan apapun dari symbol awal, sehingga produksi itu redundan (berlebih)

Contoh

terdapat tata bahasa bebas konteks :

$$S \rightarrow aSa \mid Abd \mid Bde$$

$$A \rightarrow Ada$$

$$B \rightarrow BBB \mid a$$

Bisa kita lihat :

1. Simbol variabel A tidak memiliki penurunan yang menuju terminal sehingga bisa dihilangkan.
2. Konsekuensi no (1), aturan produksi $S \rightarrow Abd$ tidak memiliki penurunan

Maka tata bahasa bebas konteks setelah disederhanakan menjadi :

$$S \rightarrow aSa \mid Bde$$

$$B \rightarrow BBB \mid a$$

Contoh, terdapat tata bahasa bebas konteks :

$$S \rightarrow Aa \mid B$$

$$A \rightarrow ab \mid D$$

$$B \rightarrow b \mid E$$

$$C \rightarrow bb$$

$$E \rightarrow aEa$$

Bisa kita lihat :

1. Aturan produksi $A \rightarrow D$, simbol variabel D tidak memiliki penurunan
2. Aturan produksi $C \rightarrow bb$, bila kita coba melakukan penurunan dari simbol awal S, dengan jalan mana pun tidak akan pernah mencapai C
3. Simbol variabel E tidak memiliki aturan produksi yang menuju terminal ($E \rightarrow aEa$ satu-satunya aturan produksi dari E)
4. Konsekuensi no (3) Aturan produksi $B \rightarrow E$, simbol variabel E tidak memiliki penurunan

maka dari tata bahasa bebas konteks di atas, produksi yang *useless* :

$$A \rightarrow D$$

$$C \rightarrow bb$$

$$E \rightarrow aEa$$

$$B \rightarrow E$$

maka tata bahasa bebas konteks setelah disederhanakan menjadi :

$$S \rightarrow Aa \mid B$$

$$A \rightarrow ab$$

$$B \rightarrow b$$

Contoh tata bahasa bebas konteks :

$$S \rightarrow aAb \mid cEB$$

$$A \rightarrow dBE \mid eeC$$

$$B \rightarrow ff$$

$C \rightarrow ae$

$D \rightarrow h$

Kita lihat :

1. Aturan produksi $S \rightarrow cEB$, $A \rightarrow dBE$ (E tidak memiliki penurunan)
2. Aturan produksi $D \rightarrow h$, redundan

Sisa aturan produksi

$S \rightarrow aAb$

$A \rightarrow eeC$

$B \rightarrow ff$

$C \rightarrow ae$

Kita lihat sekarang $B \rightarrow ff$ juga redundan sehingga hasil penyederhanaan menjadi :

$S \rightarrow aAb$

$A \rightarrow eeC$

$C \rightarrow ae$

Contoh tata bahasa bebas konteks :

$S \rightarrow aB$

$A \rightarrow bcD \mid dAC$

$B \rightarrow e \mid Ab$

$C \rightarrow bCd \mid adF \mid ab$

$F \rightarrow cFB$

1. Aturan produksi $A \rightarrow bcD$, variabel D tidak memiliki penurunan
2. Konsekuensi no (1), simbol variabel A tidak memiliki penurunan yang menuju terminal (tinggal $A \rightarrow dAC$)
3. konsekuensi no (2), $B \rightarrow Ab$ tidak memiliki penurunan
4. Simbol variabel F tidak memiliki penurunan yang menuju terminal
5. Konsekuensi no (4), $C \rightarrow adF$ tidak memiliki penurunan

Setelah disederhanakan menjadi :

$S \rightarrow aB$

$B \rightarrow e$

$C \rightarrow bCd \mid ab$

Contoh tata bahasa bebas konteks :

$S \rightarrow aBD$

$B \rightarrow cD \mid Ab$

$D \rightarrow ef$

$A \rightarrow Ed$

$F \rightarrow dc$

1. Aturan produksi $A \rightarrow Ed$, E tidak memiliki penurunan
2. Aturan produksi $F \rightarrow dc$, redundan

Sisa aturan produksi :

$$S \rightarrow aBD$$

$$B \rightarrow cD \mid Ab$$

$$D \rightarrow ef$$

Kita lihat sekarang $B \rightarrow Ab$, A tidak memiliki penurunan

Aturan produksi setelah disederhanakan :

$$S \rightarrow aBD$$

$$B \rightarrow cD$$

$$D \rightarrow ef$$

Contoh tata bahasa bebas konteks :

$$S \rightarrow Abc \mid ab$$

$$A \rightarrow AAA \mid \varepsilon$$

Aturan produksi setelah disederhanakan :

$$S \rightarrow Abc \mid ab$$

$$A \rightarrow AAA \mid \varepsilon$$

Ingat $A \rightarrow \varepsilon$ juga harus diperhitungkan

- * Prinsipnya setiap kali melakukan penyederhanaan kita periksa lagi aturan produksi yang tersisa, apakah semua produksi yang *useless* sudah dihilangkan.

Penghilangan Produksi Unit

Produksi unit adalah produksi dimana ruas kiri dan ruas kanan aturan produksi hanya berupa satu symbol variabel, misalkan $A \rightarrow B$, $C \rightarrow D$. Keberadaan produksi unit membuat tata bahasa memiliki kerumitan yang tak perlu atau menambah panjang penurunan. Penyederhanaan ini dilakukan dengan melakukan pengantian aturan produksi unit.

Contoh tata bahasa bebas konteks :

$$S \rightarrow Sb$$

$$S \rightarrow C$$

$$C \rightarrow D$$

$$C \rightarrow ef$$

$$D \rightarrow dd$$

Kita lakukan penggantian berurutan mulai dari aturan produksi yang paling dekat menuju ke penurunan terminal-terminal (\Rightarrow dibaca menjadi) :

- $C \rightarrow D \Rightarrow C \rightarrow dd$
- $S \rightarrow C \Rightarrow S \rightarrow dd \mid ef$

Sehingga aturan produksi setelah penyederhanaan :

$S \rightarrow Sb$
 $S \rightarrow dd \mid ef$
 $C \rightarrow dd$
 $C \rightarrow ef$
 $D \rightarrow dd$

Contoh tata bahasa bebas konteks :

$S \rightarrow A$
 $S \rightarrow Aa$
 $A \rightarrow B$
 $B \rightarrow C$
 $B \rightarrow b$
 $C \rightarrow D$
 $C \rightarrow ab$
 $D \rightarrow b$

Penggantian yang dilakukan :

- $C \rightarrow D \Rightarrow C \rightarrow b$
- $B \rightarrow C \Rightarrow B \rightarrow b \mid ab$, karena $B \rightarrow b$ sudah ada, maka kita cukup tuliskan $B \rightarrow ab$
- $A \rightarrow B \Rightarrow A \rightarrow ab \mid b$
- $S \rightarrow A \Rightarrow S \rightarrow ab \mid b$

Sehingga aturan produksi setelah penyederhanaan :

$S \rightarrow ab \mid b$
 $S \rightarrow Aa$
 $A \rightarrow ab \mid b$
 $B \rightarrow ab$
 $B \rightarrow b$
 $C \rightarrow b$
 $C \rightarrow ab$
 $D \rightarrow b$

Contoh tata bahasa bebas konteks :

$S \rightarrow Cba \mid D$
 $A \rightarrow bbC$
 $B \rightarrow Sc \mid ddd$
 $C \rightarrow eA \mid f \mid C$
 $D \rightarrow E \mid SABC$

$$E \rightarrow gh$$

Penggantian yang dilakukan :

- $D \rightarrow E \Rightarrow D \rightarrow GH$
- $C \rightarrow C$, kita hapus
- $S \rightarrow D \Rightarrow S \rightarrow gh \mid SABC$

Sehingga aturan produksi setelah penyederhanaan :

$$S \rightarrow Cba \mid gh \mid SABC$$

$$A \rightarrow bbC$$

$$B \rightarrow Sc \mid ddd$$

$$C \rightarrow eA \mid f$$

$$D \rightarrow gh \mid SABC$$

$$E \rightarrow gh$$

Penghilangan Produksi ε

Produksi ε adalah produksi dalam bentuk

$$\alpha \rightarrow \varepsilon$$

atau bisa dianggap sebagai produksi kosong. Penghilangan produksi ε dilakukan dengan melakukan penggantian produksi yang memuat variabel yang bisa menuju produksi ε , atau biasa disebut nullable. Prinsip pengantiannya bisa dilihat kasus berikut :

$$S \rightarrow bcAd$$

$$A \rightarrow \varepsilon$$

Pada kasus di atas A *nullable*, serta $A \rightarrow \varepsilon$ satu-satunya produksi dari A , maka variabel A bisa diabaikan, hasil penyederhanaan tata bahasa bebas konteks menjadi :

$$S \rightarrow bcd$$

Tetapi bila kasusnya :

$$S \rightarrow bcAd$$

$$A \rightarrow bd \mid \varepsilon$$

Pada kasus di atas A *nullable*, tapi $A \rightarrow \varepsilon$ bukan satu-satunya produksi dari A , maka hasil penyederhanaan :

$$S \rightarrow bcAd \mid bcd$$

$$A \rightarrow bd$$

Contoh tata bahasa bebas konteks :

$$S \rightarrow Ab \mid Cd$$

$$A \rightarrow d$$

$$C \rightarrow \varepsilon$$

Variabel yang *nullable* adalah A, B, C. Dari $S \rightarrow AB$, maka S juga *nullable*. Kita lakukan penggantian :

- $A \rightarrow aCa \Rightarrow A \rightarrow aa$
- $B \rightarrow bA \Rightarrow B \rightarrow bA \mid b$
- $B \rightarrow BB \Rightarrow B \rightarrow BB \mid B$
- $A \rightarrow abB \Rightarrow A \rightarrow abB \mid ab$
- $S \rightarrow AB \Rightarrow S \rightarrow AB \mid A \mid B \mid \varepsilon$
- $C \rightarrow \varepsilon, B \rightarrow \varepsilon, A \rightarrow \varepsilon$ dihapus

* Perhatikan : untuk penggantian $S \rightarrow AB$ kita tetap mempertahankan produksi $S \rightarrow \varepsilon$, karena S merupakan symbol awal. Ini merupakan satu-satunya pengecualian produksi ε yang tidak dihapus, yaitu produksi ε yang dihasilkan oleh symbol awal.

Hasil akhir penyederhanaan :

$$\begin{aligned} S &\rightarrow AB \mid A \mid B \mid \varepsilon \\ A &\rightarrow abB \mid ab \mid aa \\ B &\rightarrow bA \mid b \mid BB \mid B \end{aligned}$$

Contoh tata bahasa bebas konteks :

$$\begin{aligned} S &\rightarrow aAb \\ A &\rightarrow aAb \mid \varepsilon \end{aligned}$$

Hasil penyederhanaan :

$$\begin{aligned} S &\rightarrow aAb \mid ab \\ A &\rightarrow aAb \mid ab \end{aligned}$$

Contoh tata bahasa bebas konteks :

$$\begin{aligned} S &\rightarrow ABaC \\ A &\rightarrow BC \\ B &\rightarrow b \mid \varepsilon \\ C &\rightarrow D \mid \varepsilon \\ D &\rightarrow d \end{aligned}$$

Hasil penyederhanaan :

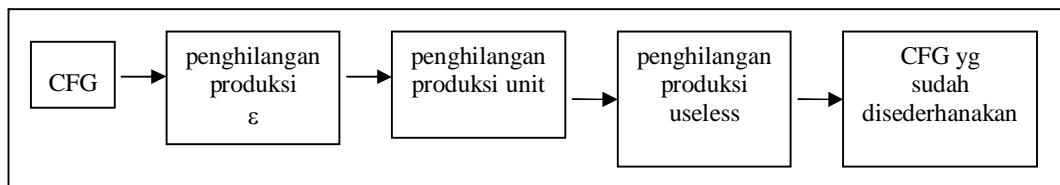
$$\begin{aligned} S &\rightarrow ABaC \mid BaC \mid AaC \mid ABa \mid aC \mid Aa \mid Ba \mid a \\ A &\rightarrow B \mid C \mid BCB \\ B &\rightarrow b \\ C &\rightarrow D \end{aligned}$$

$$D \rightarrow d$$

Pada prakteknya ketiga penyederhanaan tersebut (penghilangan *useless*, unit, ϵ) dilakukan bersama pada suatu tata bahasa bebas konteks, yang nantinya menyiapkan tata bahasa bebas konteks tersebut untuk diubah ke dalam suatu bentuk normal Chomsky yang akan dibahas pada bab selanjutnya. Hal yang memerlukan perhatian adalah penghilangan suatu tipe produksi bisa menghasilkan produksi tipe yang lain , hal ini didasari kenyataan bahwa penghilangan produksi ϵ bisa menghasilkan produksi unit. Tapi perhatikan juga bahwa penghilangan produksi unit tidak menghasilkan produksi ϵ , dan penghilangan produksi useless tidak menghasilkan produksi unit maupun produksi ϵ . Maka kita bisa menghapuskan semua produksi yang tidak diinginkan tersebut dengan melakukan urutan sebagai berikut :

1. Hilangkan produksi ϵ
2. Hilangkan produksi unit
3. Hilangkan produksi useless

Bisa dilihat alur penyederhanaan tata bahasa bebas konteks pada gambar berikut.



Gambar 1 Penyederhanaan tata bahasa bebas konteks

Hasil yang kita peroleh adalah tata bahasa yang sudah bebas dari ketiga jenis produksi tersebut. Kita harus mencoba untuk melakukan ketiga penyederhanaan tersebut pada aturan produksi berikut :

$$S \rightarrow AA \mid C \mid bd$$

$$A \rightarrow Bb \mid \epsilon$$

$$B \rightarrow AB \mid d$$

$$C \rightarrow de$$

Pertama-tama kita lakukan penghilangan produksi ϵ , sehingga aturan produksi menjadi :

$$S \rightarrow A \mid AA \mid C \mid bd$$

$$A \rightarrow Bb$$

$$B \rightarrow B \mid AB \mid d$$

$$C \rightarrow de$$

Nampak bahwa penghilangan produksi ε berpotensi untuk menghasilkan produksi unit yang baru yang sebelumnya tidak ada.

Selanjutnya kita lakukan penghilangan produksi unit menjadi :

$$S \rightarrow Bb \mid AA \mid de \mid bd$$

$$A \rightarrow Bb$$

$$B \rightarrow AB \mid d$$

$$C \rightarrow de$$

Anda lihat, penghilangan produksi unit bisa menghasilkan produksi *useless*. Terakhir kita lakukan penghilangan produksi *useless* :

$$S \rightarrow Bb \mid AA \mid de \mid bd$$

$$A \rightarrow Bb$$

$$B \rightarrow AB \mid d$$

Bisa anda periksa, hasil akhir aturan produksi tidak lagi memiliki produksi ε , produksi unit maupun produksi *useless*.